

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: May 13, 2004, 08:21:22 ; Search time 67.5 Seconds

(without alignments)
10919.267 Million cell updates/sec

Title: us-10-030-529a-1

Perfect score: 2048

Sequence: 1 ataaatcgtcattgacatt.....aagcgttaagtgaggac 1168

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5

Ygapop 10.0 , Ygapext 0.5

Fgapop 6.0 , Fgapext 7.0

Delop 6.0 , Delext 7.0

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 2034082

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlp
-Q=/cgn2 1/USPTO spoil_p/US10030529/runat 13052004 081452 2610/app query.fasta 1.1351
-DB=SPTRMBL 25 -QFMT=fastan -SUFFIX=rspt -MINMATCH=0.1 -IOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40 cdi -LIST=45
-DOCALLGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODEB=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10030529 @CGN 1.1 86 @runat 13052004 081452 2610 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phase.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rv.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1342	65.5	257	16	Q9K2H6

Q9K2H6 haemophilus

2	1299.5	63.5	264	2	Q9K3B8	Q9k3b8	haemophilus	
3	1289	62.9	271	2	Q9L916	Q9l916	haemophilus	
4	1283	62.6	273	2	Q9L917	Q9l917	haemophilus	
5	1262	61.6	248	2	Q9L919	Q9l919	haemophilus	
6	1205	58.8	263	2	Q9L918	Q9l918	haemophilus	
C	7	235	11.8	440	16	Q7VN22	Q7vn22	haemophilus
8	225	11.0	295	2	Q8RR69	Q8rr69	actinobacil	
9	225	11.0	295	2	Q8RO63	Q8rq63	actinobacil	
10	222	10.8	504	2	Q9LA56	Q9la56	escherichia	
11	222	10.8	511	9	Q9MC18	Q9mc18	bacteriopha	
12	220	10.7	392	2	Q9LA60	Q9la60	escherichia	
13	217	10.6	355	2	Q8KUJ1	Q8kuj1	neisseria m	
14	217	10.6	362	2	Q8KH7	Q8kh7	neisseria m	
15	217	10.6	364	16	Q9JXK7	Q9jxk7	neisseria m	
16	217	10.6	459	2	Q8VW24	Q8vw24	escherichia	
17	209	10.2	391	2	Q8KUJ6	Q8kuj6	neisseria m	
18	209	10.2	398	2	Q8KI42	Q8ki42	neisseria m	
19	208.5	10.2	487	2	Q9LA53	Q9la53	escherichia	
20	204	10.0	405	2	Q8KHP5	Q8khp5	neisseria m	
21	204	10.0	405	2	Q8KH85	Q8kh85	neisseria m	
22	198	9.7	668	2	Q8GH86	Q8gh86	moraxella c	
23	196	9.6	630	2	Q9XD55	Q9xd55	moraxella c	
24	193	9.4	613	2	Q9XD53	Q9xd53	moraxella c	
25	193	9.4	616	2	Q848S2	Q848s2	moraxella c	
26	193	9.4	877	2	Q848S1	Q848s1	moraxella c	
27	193	9.4	889	2	Q9L961	Q9l961	moraxella c	
28	193	9.4	894	2	Q9L962	Q9l962	moraxella c	
29	190.5	9.3	576	2	Q54407	Q54407	moraxella c	
30	190.5	9.3	674	2	Q9XD51	Q9xd51	moraxella c	
31	190.5	9.3	684	2	Q9L963	Q9l963	moraxella c	
32	190.5	9.3	686	2	Q8RTB2	Q8rtb2	moraxella c	
33	186	9.1	422	2	Q93KR4	Q93kr4	versinia en	
34	186	9.1	422	2	Q56930	Q56930	versinia en	
35	180.5	8.8	422	2	Q84GR6	Q84gr6	versinia en	
36	161	7.9	454	2	Q85267	Q85267	versinia en	
37	159.5	7.8	2712	16	Q9F3X5	Q9f3x5	pasteurella	
C	38	151	7.6	419	16	Q9CLN4	Q9cln4	pasteurella
39	132	6.4	155	16	Q8YJ81	Q8yj81	brucella me	
40	130	6.3	278	16	Q8FYM1	Q8fym1	brucella su	
C	41	130	6.5	420	16	Q8FEZ1	Q8fez1	escherichia
42	129	6.5	227	16	Q7UBW0	Q7ubw0	shigella fl	
C	43	129	6.5	399	16	Q83K08	Q83k08	shigella fl
C	44	129	6.5	413	16	Q8XFY3	Q8xfy3	salmonella
C	45	129	6.5	428	16	Q8X9C3	Q8x9c3	escherichia

ALIGNMENTS

RESULT 1

Q9K2H6	Q9K2H6	PRELIMINARY;	PRT;	257 AA.
AC	Q9K2H6;			
DT	01-OCT-2000 (TrEMBLrel. 15, Created)			
DT	01-OCT-2000 (TrEMBLrel. 15, Last sequence update)			
DT	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)			
DE	Serum resistance protein DsrA (Serum resistance protein DsrA).			
GN	DSRA OR HD0769.			
OS	Haemophilus ducreyi.			
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;			
OC	Pasteurellaceae; Haemophilus.			
OX	NCBI_TaxID=730;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=CIP542(CDC), 35000, and 406;			
RX	MEDLINE=20143779; PubMed=10678980;			
RA	Elkins C., Morrow K.J. Jr., Olsen B.;			
RT	"Serum Resistance in Haemophilus ducreyi Requires Outer Membrane			
RL	Protein DsrA.";			
RL	Infect. Immun. 68:1608-1619(2000).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=35000HP / ATCC 700724;			
RA	Munson R.S. Jr., Ray W.C., Mahairas G., Sabo P., Mungur R.,			

RA Johnson L., Nguyen D., Wang J., Forst C., Hood L.;
RT "The complete genome sequence of Haemophilus ducreyi."
KL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF197007; AAF37813.1; -
DR EMBL; AF187001; AAF37807.1; -
DR EMBL; AF187005; AAF37811.1; -
DR EMBL; AE017152; AAP95674.1; -
DR InterPro; IPR005594; Yada.
DR Pfam; PF03895; Yada; 1.
KW Complete proteome.
SQ SEQUENCE 257 AA; 28524 MW; F497BF2CD5666938 CRC64;

Alignment Scores:
Pred. No.: 2.93e-104 Length: 257
Score: 1342.00 Matches: 257
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 65.53% Indels: 0
DB: 16 Gaps: 0

US-10-030-529A-1 (1-1168) x Q9K2H6 (1-257)

QY 101 ATGAAATTAATGTTTGTAGTCCGCTAGTGGATTAGCTTCTTCTACTATTACAACTG 160
DB 1 MetLysileLysCysLeuValAlaValValGlyLeuAlaCysSerThrIleThrMet 20
QY 161 GCTCAGCAGCGCCAAAGTTTCTGGAGTATCTTTGTATAGTATGACTATGACTAT 220
DB 21 AlaGlnGlnProProLysPheAlaGlyValSerSerLeuTyrSerTyrGluTyrAspTyr 40
QY 221 GGTAAAGGTAATGACATGCTCTAATAAGAGCGGTTTCGATATTAAAGTCCAGGATT 280
DB 41 GlyLysGlyLysTyrThrTyrSerAsnGluGlyPheAspIleLysValProGlyIle 60
QY 281 AAATGAAGCCAAAGAAAGTATCTTAAACAGGCTACTTATCTTGAATTACAGATTAT 340
DB 61 LysMetLysProLysGluTyrPheSerLysGlnAlaThrTyrLeuGluLeuGlnHisTyr 80
QY 341 ATGCTTATATCTCTGCTGACATATCTCTCGGCTTCTCTAGCCCTACTATG 400
DB 81 MetProTyrThrProValLeuValThrTyrAlaProGlyValSerProSerProIleLeu 100
QY 401 TTATATCCGATGCTCTGATCAACTTGAATTAATCGCAGCAGCTGAAATTGAAAT 460
DB 101 LeuTyrProMetSerAspProAspGlnLeuGlyIleAsnArgGlnGlnLeuLysLeuAsn 120
QY 461 TTGTATAGTATTTTAAACGATTTAAGACACGATTTTAAATTAAGTTCTTGATGACGT 520
DB 121 LeuTyrSerTyrPheAsnAspLeuArgHisAspPheLysLeuLysValLeuAspAlaArg 140
QY 521 ATTTCCAAATTAACAAATATTGATATAGTAAATTTTACTAGACTCGGTACT 580
DB 141 IleSerLysAsnLysGlnAsnIleAspThrIleSerLysTyrLeuLeuGluLeuGlyThr 160
QY 581 TATTATAGTATTTTAAACGATTTTAAATTAAGTTCTTGATGACGT 640
DB 161 TyrLeuAspAspSerTyrArgMetMetGluGlnAsnThrHisAsnIleAsnLysLeuSer 180
QY 641 AAAGAAATTCGAATGTTTACCAACCAATCAGCATTTCTATGTTAGTGCACCAAT 700
DB 181 LysGluLeuGlnThrGlyLeuAlaAsnGlnSerAlaLeuSerMetMetValGlnProAsn 200
QY 701 GGTGTAGCAAAACGAGGTTTCTGCTGGGTAGGAGGTTATAGATAAACTGCATTA 760
DB 201 GlyValGlyLysThrSerValSerAlaAlaValGlyTyrArgAspLysThrAlaLeu 220
QY 761 GCCATTGGTCTCGGCTCAGCATTTACTGATCCGCTTTTACCGCTTAAAGCGGTGACGGTTC 820
DB 221 AlalleGlyValGlySerArgIleThrAspArgPheThrAlaLysAlaGlyValAlaPhe 240
QY 821 AATACCTACAATGGCGGATGCTTATGCTGCTTCTGTTGGTTATGAATTC 871
DB 241 AsnThrTyrAsnGlyGlyMetSerTyrGlyAlaSerValGlyTyrGluPhe 257

RESULT 2
Q9K3B8 PRELIMINARY; PRT; 264 AA.
AC Q9K3B8
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 22, Last annotation update)
DE Serum resistance protein DsrA.
GN DSRP.
OS Haemophilus ducreyi.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=730;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CIPA77, and CIPA75;
RX MEDLINE=20143779; PubMed=10678980;
RA Elkins C., Morrow K.J. Jr., Olsen B.;
RT "Serum Resistance in Haemophilus ducreyi Requires Outer Membrane
Protein DsrA";
RL Infect. Immun. 68:1608-1619 (2000).
DR EMBL; AF187003; AAF37809.1; -
DR EMBL; AF187002; AAF37808.1; -
DR InterPro; IPR005594; Yada.
DR Pfam; PF03895; Yada; 1.
SQ SEQUENCE 264 AA; 29444 MW; 11AF6C124D7ED126 CRC64;

Alignment Scores:
Pred. No.: 1.09e-100 Length: 264
Score: 1299.50 Matches: 253
Percent Similarity: 95.83% Conservative: 0
Best Local Similarity: 95.83% Mismatches: 4
Query Match: 63.45% Indels: 7
DB: 2 Gaps: 1

US-10-030-529A-1 (1-1168) x Q9K3B8 (1-264)

QY 101 ATGAAATTAATGTTTGTAGTTCGCTAGTGGATTAGCTTCTTCTACTATTACAACTG 160
DB 1 MetLysileLysCysLeuValAlaValValGlyLeuAlaCysSerThrIleThrMet 20
QY 161 GCTCAGCAGCGCCAAAGTTTCTGGAGTATCTTCTTGTATAGCTATGACTATG 220
DB 21 AlaGlnGlnProProLysPheAlaGlyValSerSerLeuTyrSerTyrGluTyrAspTyr 40
QY 221 GGTAAAGGTAATGACATGCTGCTAATAAGAGCGGTTTCGATATTAAAGTCCAGGATT 280
DB 41 GlyLysGlyLysTyrThrTyrSerAsnGluGlyGlyPheAspIleLysValProGlyIle 60
QY 281 AAATGAAGCCAAAGAAAGTATCTTAAACAGGCTACTTATCTTGAATTACAGCATTA 340
DB 61 LysMetLysProLysGluTyrPheSerLysGlnAlaThrTyrLeuGluLeuGlnHisTyr 80
QY 341 ATGCTTATATCTCTGCTGACATATCTCTCGGCTTCTCTAGCCCTACTATG 400
DB 81 MetProTyrThrProValLeuValThrTyrAlaHisAspValProProSerSerIleLeu 100
QY 401 TTATATCCGATGCTCTGATCAACTTGGATTAATTCGCGACAGCTGAAATTGAAAT 460
DB 101 LeuTyrProMetSerAspProAspGlnLeuGlyIleAsnArgGlnGlnLeuLysLeuAsn 120
QY 461 TTGTATAGTATTTTAAACGATTTTAAATTAAGTTCTTGATGACGT 520
DB 121 LeuTyrSerTyrPheAsnAspLeuArgHisAspPheLysLeuLysValLeuAspAlaArg 140
QY 521 ATTTCCAAATTAACAAATATTGATATAGTAAATTTTACTAGAACTCGGTACT 580
DB 141 IleSerLysAsnLysGlnAsnIleAspThrIleSerLysTyrLeuLeuGluLeuGlyThr 160
QY 581 TATTATAGTATTTTAAACGATTTTAAATTAAGTTCTTGATGACGT 619
DB 161 TyrLeuAspAspSerTyrArgMetMetGluGlnAsnThrHisAsnIleAsnLysAsnThr 180

```

Oy 620 CATATATCAATAAGTTCTCTAAAGAAATTGCAAACTGGTTTAGCCAAACCAATCAGCATG 679
Db 181 HisanilleAsnLysLeuSerLysGluLeuGlnThrGlyLeuAlaAsnGlnSerAlaLeu 200
Oy 680 TCTATGTTAGTGCACCAAAATGGTGTAGGCAAAACGAGCGTTCTCTGCTGCGGTAGGAGGT 739
Db 201 SerMetLeuValGlnProAsnGlyValGlyLysThrSerValSerAlaAlaValGlyGly 220
Oy 740 TATAGATATAAAGTGCATTAGCCATTGGTGTGGCTCACCACCATCTAGTCGCTTTACC 799
Db 221 TyrArgAspLysThrAlaLeuAlaIleGlyValGlySerArgIleThrAspArgPheThr 240
Oy 800 GCTAAAGCGGTGTAGCGTTCAATACCTACCAATGGCGCATGCTCTATGGTGCTTCGTGTT 859
Db 241 AlalysAlaGlyValAlaPheAsnThrTyrAsnGlyGlyMetSerTyrGlyAlaSerVal 260
Oy 860 GGTTATGAATTC 871
Db 261 GlyTyrGluPhe 264

RESULT 3
Q9L916
ID Q9L916 PRELIMINARY; PRT; 271 AA.
AC Q9L916;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Serum resistance protein DsrA.
GN DSR.
OS Haemophilus ducreyi.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=730;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=M90-02;
RX MEDLINE=20143779; PubMed=10678980;
RA Elkins C., Morrow K.J. Jr., Olsen B.;
RT "Serum Resistance in Haemophilus ducreyi Requires Outer Membrane Protein DsrA.";
RL Infect. Immun. 68:1608-1619 (2000).
DR EMBL; AF187009; AAF37815.1; -.
DR InterPro; IPR005594; Yada.
DR Pfam; PF03895; Yada; 1.
SQ SEQUENCE 271 AA; 30091 MW; 5A9DC55C6DA6EF1C CRC64;

Alignment Scores:
Pred. No.: 8.28e-100 Length: 271
Score: 1289.00 Matches: 252
Percent Similarity: 92.99% Conservative: 0
Best Local Similarity: 92.99% Mismatches: 5
Query Match: 62.94% Indels: 14
DB: 2 Gaps: 1

US-10-030-529a-1 (1-1168) x Q9L916 (1-271)
Oy 101 ATGAAAATTAAATGTTAGTTCGCGTAGTGGATTAGCTTGTTCTACTATTACAAACATG 160
Db 1 MetLysIleLysCysLeuValAlaValAlaValGlyLeuAlaCysSerThrIleThrMet 20
Oy 161 GCTCAGACGCCCAAGATTGCTGGAGTATCTCTTTGTATAGCTATGACTATGACTAT 220
Db 21 AlaGlnGlnProLysPheAlaGlyValSerSerLeuTyrSerTyrGluTyrAspTyr 40
Oy 221 GGTAAAGGTAATGACTTGTCTAATGAAGCGGTTTCGATATTAAAGTCCAGGCGATT 280
Db 41 GlyLysGlyLysIleThrTrpSerAsnGluGlyGlyPheAspIleLysValProGlyIle 60
Oy 281 AAAATGAAGCAAAAGAAATGATTCTTAAACAGCGCTACTTATCTTGAATACACATAT 340
Db 61 LysMetLysProLysGluTrpIleSerLysGlnAlaThrTyrLeuGluLeuGlnHisTyr 80

```

```

Oy 341 ATGCTTATATCTCTGTTCTGTCGACATATGCTCTCGCGGTTTCTCTAGCCCTATACTG 400
Db 81 MetProTyrThrProValLeuValThrSerAlaProAspValProProSerSerIleLeu 100
Oy 401 TTATATCCGATGTCGATCTGATCAACTTGGATAAATCGGCAGCAGCTGAAATTGAAT 460
Db 101 LeuTyrProMetSerAspProAspGlnLeuGlyIleAsnArgGlnGlnLeuLysLeuAsn 120
Oy 461 TTGTATAGTTATTTTAACGATTAAAGACAGATTTTAAATTAAGTCTTGATGCAGGT 520
Db 121 LeuTyrSerTyrPheAsnAspLeuArgHisAspPheLysLeuLysValLeuAspAlaArg 140
Oy 521 ATTTCCAAAAATAACCAAAATATTGATCACTATAAGTAAATATTACTAGAACTGGGTACT 580
Db 141 IleSerLysAsnLysGlnAsnIleAspThrIleSerLysTyrLeuLeuGluLeuGlyThr 160
Oy 581 TATTAGATGATCTTATCGTATGATGAACAA----- 613
Db 161 TyrLeuAspGlySerTyrArgMetMetGluGlnAsnThrHisAsnIleAsnLysAsnThr 180
Oy 614 -----AATACACATAATCAATAAGTTGTCTAAAGATTGCTCAAACTGGT 658
Db 181 HisAsnIleAsnLysAsnThrHisAsnIleAsnLysLeuSerLysGluLeuGlnThrGly 200
Oy 659 TTAGCCAAACCAATCAGCATTTCTATGTTAGTGCACCAAAATGTTGTAGGCAAAACGAGC 718
Db 201 LeuAlaAsnGlnSerAlaLeuSerMetLeuValGlnProAsnGlyValGlyLysThrSer 220
Oy 719 GTTCTCTGCGGTAGGAGTTATAGAGATAAACTGCATTAGCCATTGGTGTGCGGTCA 778
Db 221 ValSerAlaAlaValGlyGlyTyrArgAspLysThrAlaLeuAlaIleGlyValGlySer 240
Oy 779 CGCATTTACTGATCGCTTTACCGCTAAAGCGGTGTAGCGTTCAATACCTACATGGCGGC 838
Db 241 ArgIleThrAspArgPheThrAlaLysAlaGlyValAlaPheAsnThrTyrAsnGlyGly 260
Oy 839 ATGCTTATGCTCTCTGTTGTTGTTATGAATTC 871
Db 261 MetSerTyrGlyAlaSerValGlyTyrGluPhe 271

RESULT 4
Q9L917
ID Q9L917 PRELIMINARY; PRT; 273 AA.
AC Q9L917;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Serum resistance protein DsrA.
GN DSR.
OS Haemophilus ducreyi.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=730;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=M90-02;
RX MEDLINE=20143779; PubMed=10678980;
RA Elkins C., Morrow K.J. Jr., Olsen B.;
RT "Serum Resistance in Haemophilus ducreyi Requires Outer Membrane Protein DsrA.";
RL Infect. Immun. 68:1608-1619 (2000).
DR EMBL; AF187008; AAF37814.1; -.
DR InterPro; IPR005594; Yada.
DR Pfam; PF03895; Yada; 1.
SQ SEQUENCE 273 AA; 30281 MW; 8C620D6B5544E3EE CRC64;

Alignment Scores:
Pred. No.: 2.64e-99 Length: 273
Score: 1283.00 Matches: 253
Percent Similarity: 92.67% Conservative: 0
Best Local Similarity: 92.67% Mismatches: 4
Query Match: 62.65% Indels: 16
DB: 2 Gaps: 2

```

```
US-10-030-529A-1 (1-1168) x Q9L917 (1-273)
QY 101 ATGAAATTAATAAGTTTGTAGTCCGTAGTGGGATTAGCTTCTTCTACTATTACAAATG 160
Db 1 MetLysIleLysCysLeuValAlaValGlyLeuAlaCysSerThrIleThrMet 20
QY 161 GCTCAGCAGCGCCAAAGTTGCTCGAGTATCTCTTTGTATAGCTATGAGTATGACTAT 220
Db 21 AlaGlnGlnProProLysPheAlaGlyValSerSerLeuTyrSerTyrGluTyrAspTyr 40
QY 221 GGTAAAGGTAATAGTACTTGCTTAATCAAGCGGTTTCGATATTAAAGTCCAGGGATT 280
Db 41 GlyLysGlyLysTyrThrTyrSerAsnGluGlyGlyPheAspIleLysValProGlyIle 60
QY 281 AAATGAAGCAAAAGATGATTTCTAAACAGGCTACTTATCTTGAATTTACAGCATAT 340
Db 61 LysMetLysProLysGluTyrPheAsnAspLeuArgHisAspPheLysLeuLysValLysAsp 80
QY 341 ATGCTTATATCTCTGTTCTGTGACATATGCTCTGGCGTTTCCTCCTAGC-----CCT 394
Db 81 MetProTyrThrProValLeuValThrSerAlaProAspValSerProSerSerIleSer 100
QY 395 ATACTGTTATATCCGATGCTCTGATCCTGATCAACTTGAATAAATCGGCAGCAGCTGAAA 454
Db 101 IleLeuLeuTyrProMetSerAspProAspGlnLeuGlyIleAsnArgGlnGlnLeuLys 120
QY 455 TTGAATTTGTATAGTATTTTAAACGATTTAGACAGCATTTTAAATTTAAAGTTCTTGAT 514
Db 121 LeuAsnLeuTyrSerTyrPheAsnAspLeuArgHisAspPheLysLeuLysValLysAsp 140
QY 515 GCACGTATTTCCAAAATAAACAATAATATGATATATGATATGATATGATATGATATGAT 574
Db 141 AlaArgIleSerLysAsnLysGlnAsnIleAspThrIleSerLysTyrLeuLeuGluLeu 160
QY 575 GGTACTTATTTAGATGTTCTTATCGTATGATGGAACAA-----613
Db 161 GlyThrTyrLeuAspGlySerTyrArgMetMetGluGlnAsnThrHisAsnIleAsnLys 180
QY 614 -----AATACACATAATATCAATAGTTGCTTAAAGAATTGCAA 652
Db 181 AsnThrHisAsnIleAsnLysAsnThrHisAsnIleAsnLysLeuSerLysGluLeuGln 200
QY 653 ACTGGTTTAGCCAAACAATCAGCATGCTATGTTAGTGAACCAAAATGGTGTAGGCAAA 712
Db 201 ThrGlyLeuAlaAsnGlnSerAlaLeuSerMetLeuValGlnProAsnGlyValGlyLys 220
QY 713 ACGAGCGTTTCTGCTCGGTAGAGGTTATAGAGATAAAACTGCAATTAGCCATTGGTGTC 772
Db 221 ThrSerValSerAlaAlaValGlyGlyTyrArgAspLysThrAlaLeuAlaIleGlyVal 240
QY 773 GCCTCAGCATTTACTGATCGCTTTACCGCTAAAGCGGTGATGCGTTCAATACCTACAAAT 832
Db 241 GlySerArgIleThrAspArgPheThrAlaLysAlaGlyValAlaPheAsnThrTyrAsn 260
QY 833 GSCGCGATGCTTATGGTCTTCTGTTGTTATGAATTC 871
Db 261 GlyGlyMetSerTyrGlyAlaSerValGlyTyrGluPhe 273

RESULT 5
Q9L919
ID Q9L919 PRELIMINARY; PRT; 248 AA.
AC Q9L919,
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Serum resistance protein DsrA.
GN DsrA.
OS Haemophilus ducreyi.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=730;
RN (1)
```

```
RP SEQUENCE FROM N.A.
RC STRAIN=CIP542(Can); PubMed=10678980;
RX MEDLINE=20143779; Morrow K.J. Jr., Olsen B.;
RA Elkins C., "Serum Resistance in Haemophilus ducreyi Requires Outer Membrane
RT Protein DsrA.";
RL Infect. Immun. 68:1608-1619(2000).
DR EMBL; AF187004; AAF37810.1; -.
DR InterPro; IPR005594; YadaA.
DR Pfam; PF03895; YadaA; 1.
SQ SEQUENCE 248 AA; 27815 MW; 245C809F24E6A815 CRC64;

Alignment Scores:
Pred. No.: 1.52e-97 Length: 248
Score: 1262.00 Matches: 242
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.59% Mismatches: 0
Query Match: 61.62% Indels: 0
DB: 2 Gaps: 0

US-10-030-529A-1 (1-1168) x Q9L919 (1-248)
QY 101 ATGAAATTAATAAGTTTGTAGTCCGTAGTGGGATTAGCTTCTTCTACTATTACAAATG 160
Db 1 MetLysIleLysCysLeuValAlaValGlyLeuAlaCysSerThrIleThrMet 20
QY 161 GCTCAGCAGCGCCAAAGTTTGTGAGTATCTCTTTGTATAGCTATGAGTATGACTAT 220
Db 21 AlaGlnGlnProProLysPheAlaGlyValSerSerLeuTyrSerTyrGluTyrAspTyr 40
QY 221 GGTAAAGGTAATAGTACTTGCTTAATCAAGCGGTTTCGATATTAAAGTCCAGGGATT 280
Db 41 GlyLysGlyLysTyrThrTyrSerAsnGluGlyGlyPheAspIleLysValProGlyIle 60
QY 281 AAATGAAGCAAAAGATGATTTCTAAACAGGCTACTTATCTTGAATTTACAGCATAT 340
Db 61 LysMetLysProLysGluTyrPheAsnAspLeuArgHisAspPheLysLeuLysValLysAsp 80
QY 341 ATGCTTATATCTCTGTTCTGTGACATATGCTCTGGCGTTTCCTCCTAGCCTTACTG 400
Db 81 MetProTyrThrProValLeuValThrTyrAlaProGlyValSerProSerProIleLeu 100
QY 401 TTATATCCGATGCTCTGATCAACTTGAATAAATCGGCAGCAGCTGAAATTTGAAT 460
Db 101 LeuTyrProMetSerAspProAspGlnLeuGlyIleAsnArgGlnGlnLeuLysLeuAsn 120
QY 461 TTGTATAGTTTATTTAAAGATTTAAGACACAGATTTTAAATTTAAAGTTCTTGATCCAGT 520
Db 121 LeuTyrSerTyrPheAsnAspLeuArgHisAspPheLysLeuLysValLysAspAlaArg 140
QY 521 ATTTCCAAAATAACAAAATATTGATATAAGTAATAATTTACTAGAACTGGGTACT 580
Db 141 IleSerLysAsnLysGlnAsnIleAspThrIleSerLysTyrLeuLeuGluLeuGlyThr 160
QY 581 TATTTAGATGATTTCTTATCGTATGATGGAACAAAATACACATAATATCAATAAGTTGTCT 640
Db 161 TyrLeuAspAspSerTyrArgMetMetGluGlnAsnThrHisAsnIleAsnLysLeuSer 180
QY 641 AAAGAATTGCAAACTGGTTTAGCCAAACCAATCAGCATTTGCTATGTTAGTGCAACCAAT 700
Db 181 LysGluLeuGlnThrGlyLeuAlaAsnGlnSerAlaLeuSerMetLeuValGlnProAsn 200
QY 701 GGTGTAGCAAAACGAGCGTTTCTGCTGCGGTAGGAGTTATAGAGATAAATCTCATTA 760
Db 201 GlyValGlyLysThrSerValSerAlaAlaValGlyGlyTyrArgAspLysThrAlaLeu 220
QY 761 GCCATTGTTGCTCGGCTCAGCATTTACTGATCGCTTTACCCTTAAAGCGGTTAGCGTTTC 820
Db 221 AlaIleGlyValGlySerArgIleThrAspArgPheThrAlaLysAlaGlyValAlaPhe 240
QY 821 AATACCTTAC 829
Db 241 AsnThrPhe 243
```



```

RESULT 6
Q9L918      PRELIMINARY;      PRT;      263 AA.
ID   Q9L918
AC   Q9L918;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DI   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DR   Serum resistance protein Dera.
DS   DERA.
GN   Haemophilus ducreyi.
OS   Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC   Pasteurellaceae; Haemophilus.
OX   NCBI_TaxID=730;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CHIA;
RX   MEDLINE=20143779; PubMed=10678980;
RA   Elkins C., Morrow K.J. Jr., Olsen B.;
RT   "Serum Resistance in Haemophilus ducreyi Requires Outer Membrane
    Protein Dera.";
RL   Infect. Immun. 68:1608-1619 (2000).
DR   EMBL; AF187006; AAF37812.1; -.
DR   InterPro; IPR005594; Yada.
DR   Pfam; PF03895; Yada; 1.
SQ   SEQUENCE 263 AA; 29170 MW; 858B91BF765BE8F1 CRC64;

Alignment Scores:
Pred. No.:      9.32e-93      Length:      263
Score:          1205.00      Matches:    235
Percent Similarity: 92.05%      Conservative: 8
Best Local Similarity: 89.02%      Mismatches: 13
Query Match:    58.84%      Indels:     8
DB:              2          Gaps:         2

US-10-030-529A-1 (1-1168) x Q9L918 (1-263)
QY 101 ATGAAATTAATGTTAGTTCGGTAGTGGATTAGCTTCTTCTACTATTACAAATG 160
Db 1 MetLysIleuYsCysLeuValAlaValGlyLeuAlaCysSerThrIleThrMet 20
QY 161 GCTCAGCAGCGCCAAAGTTTGTGGAGTACTCTTTGTATAGCTATGATGACTAT 220
Db 21 AlaGlnGlnProProlYsPheAlaGlyValSerSerLeuAspSerTyrgluTyAspTy 40
QY 221 GGTAAAGGTAATGAGCTGTGCTTAATGAAGCGGTTTCGATATATAAGTCCAGGATT 280
Db 41 GlyLysGlyLysTrpThrTrpSerGluLysAspGlyPheAspIleLysAlaProGlyIle 60
QY 281 AAAATGAAGCCAAAGATGATTTCTAAACAGGCTACTTATCTTGATTAACAGATTAT 340
Db 61 LysMetLysProlYsLysTrpIleSerArgGlnAlaThrTyrgluTyLeuGlnHisTy 80
QY 341 ATGCTTATATCTCTGTTCTCGTCGATATATCTCTCGCGTTCCTCTAGCCCTATCTG 400
Db 81 MetProTyThrProValLeuValThrTyAla---SerAlaGluProAsnThrValLeu 99
QY 401 TTATATCCGATGCTGATCTGATCACTTGAATAAATCGGCAGCAGCTGAAATTTGAAT 460
Db 100 LeuTyProMetProAspProAspGlnLeuGlyLeuAsnArgGlnGlnLeuLysLeuAsn 119
QY 461 TTGTATAGTTATTTAAGCATTTAAGACACGATTTTAAATTAAGTTCTTGATGACGAT 520
Db 120 LeuTySerTyPheAsnAspLeuArgHisGlyPheLysLeuAsnValLeuAspAlaArg 139
QY 521 ATTTCCAAATAACAAATATTGATCTACTAATAATTTTACTAGACTGGGTACT 580
Db 140 IleSerGlnAsnLysGlnAsnIleAspThrIleSerGluTyrgluTyLeuLysLeuGlyThr 159
QY 581 TATTATAGATCTTCTATCGTATGATGAACAA-----AATACA 619
Db 160 TyrlleuaspSerSerTyArgMetMetGluGlnAsnThrHisAsnIleAsnLysAsnThr 179

RESULT 7
Q7VN22      PRELIMINARY;      PRT;      440 AA.
ID   Q7VN22
AC   Q7VN22;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DI   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DE   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DR   Hypothetical protein.
DS   Hypothetical protein.
GN   Haemophilus ducreyi.
OS   Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC   Pasteurellaceae; Haemophilus.
OX   NCBI_TaxID=730;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=35000HP / ATCC 700724;
RA   Munson R.S. Jr., Ray W.C., Mahairas G., Sabo P., Mungur R.,
    Johnson L., Nguyen D., Wang J., Forst C., Hood L.;
RT   "The complete genome sequence of Haemophilus ducreyi.";
RL   Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
DR   EMBL; AE017152; AAP95675.1; -.
SQ   SEQUENCE 440 AA; 49472 MW; AF37BC89E59ABABA CRC64;

Alignment Scores:
Pred. No.:      2.67e-11      Length:      440
Score:          235.00      Matches:    48
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:    11.78%      Indels:     0
DB:              16          Gaps:         0

US-10-030-529A-1 (1-1168) x Q7VN22 (1-440)
QY 1168 GTCGCGACCTTTACGGCTTAATTTTAAACATTTAGAAAATTTCCGATAGAAGTACC 1109
Db 393 ValArgThrPheAsnGlyLeuIleLeuGluHisLeuGluLysIleProIleGluGlyThr 412
QY 1108 CAATTTAGGCTAAACCACTTAATATTATACCGTTTTAGAGTGGCGGATTAATATGTTGAAA 1049
Db 413 GlnPheArgLeuAsnGlnLeuAsnIleThrValLeuGluValAlaAspAsnMetValLys 432
QY 1048 AAGTGAAGTAGAATGATTGAA 1025
Db 433 LysValLysValGluLeuIleGlu 440

RESULT 8
Q8RR69      PRELIMINARY;      PRT;      295 AA.
ID   Q8RR69
AC   Q8RR69;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DI   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DE   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
```



```

DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Immunoglobulin-binding protein EibC.
GN EibC.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ECOR-9;
RX MEDLINE=20187497; PubMed=10722621;
RA Sandt C.H., Hill C.W.;
RT "Four different genes responsible for nonimmune immunoglobulin-binding
RT activities within a single strain of Escherichia coli.";
RL Infect. Immun. 68:2205-2214 (2000).
DR EMBL; AF151674; AAF63035.1; -
DR InterPro; IPR008640; Hep_Hag.
DR InterPro; IPR005594; YadA.
DR Pfam; PF05658; Hep_Hag; 4.
DR Pfam; PF03895; YadA; 1.
SQ SEQUENCE 504 AA; 53158 MW; 6B5E192DD4B2771B CRC64;

Alignment Scores:
Pred. No.: 3.32e-10 Length: 504
Score: 222.00 Matches: 49
Percent Similarity: 58.14% Conservative: 26
Best Local Similarity: 37.98% Mismatches: 52
Query Match: 10.84% Indels: 2
DB: 2

US-10-030-529A-1 (1-1168) x Q9LA56 (1-504)
Qy 485 AGACAGATTTTAAATTTAAAGTTCTTGATGCAGTATTTCACAAATAAACAATAATT 544
Db 378 ArgileAspTyrAlaValGlyAlaIleAspGlyArgValThrArgAsnThrGlnSerIle 397
Qy 545 GATACATAAGTAATATTACTAGAACTGGGTACTTATTAGATGATCTTATCGTATG 604
Db 398 GluYasnSerLysAlaIleAlaAlaAsnThrArgThrLeuGlnGlnHisSerAlaArg 417
Qy 605 ATGGAACAAAATACACATAATATCAATAAGTTGCTAAAGATTGCAAACTGGTTAGCC 664
Db 418 LeuAspSerGlnArgGlnIleAsnGluAsnHisLysGluMetLysArgAlaAla 437
Qy 665 AACCAATCAGATTGCTATTGTAGTCAACCAATCGTGTAGCAAAACGAGCGTTTCT 724
Db 438 GlnSerAlaAlaLeuThrGlyLeuPheGlnProTyrSerValGlyLysPheAsnAlaThr 457
Qy 725 GCTCGGTAGGAGTTATAGATAAACTGCATTAGCCATTGGTGTGGCTACGCATT 784
Db 458 AlaAlaValGlyGlyTyrSerAspGlnAlaLeuAlaValGlyValGlyTyrArgPhe 477
Qy 785 ACTGATCGTTTACCGTTAAAGCGGTGTAGCTTCAATACCTACATGGCGGCATCTCT 844
Db 478 AsnGluGlnThrAlaAlaLysAlaGlyValAlaPheSer-----AspGlyAspAlaSer 495
Qy 845 TATGGTCTCTGTGGTTATGAATTC 871
Db 496 TrpAsnValGlyValAsnPheGluPhe 504

RESULT 11
Q9MC18 PRELIMINARY; PRT; 511 AA.
AC Q9MC18;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Immunoglobulin-binding protein EibD.
GN EibD.
OS Bacteriophage P-EibD.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae.
OX NCBI_TaxID=120163;

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ECOR-9;
RX MEDLINE=20187497; PubMed=10722621;
RA Sandt C.H., Hill C.W.;
RT "Four different genes responsible for nonimmune immunoglobulin-binding
RT activities within a single strain of Escherichia coli.";
RL Infect. Immun. 68:2205-2214 (2000).
DR EMBL; AF151675; AAF63040.1; -
DR InterPro; IPR008640; Hep_Hag.
DR InterPro; IPR005594; YadA.
DR Pfam; PF05658; Hep_Hag; 3.
DR Pfam; PF03895; YadA; 1.
SQ SEQUENCE 511 AA; 53843 MW; 061ABC75777CEA86 CRC64;

Alignment Scores:
Pred. No.: 3.33e-10 Length: 511
Score: 222.00 Matches: 49
Percent Similarity: 58.14% Conservative: 26
Best Local Similarity: 37.98% Mismatches: 52
Query Match: 10.84% Indels: 2
DB: 1

US-10-030-529A-1 (1-1168) x Q9MC18 (1-511)
Qy 485 AGACAGATTTTAAATTTAAAGTTCTTGATGCAGTATTTCACAAATAAACAATAATT 544
Db 385 ArgileAspTyrAlaValGlyAlaIleAspGlyArgValThrArgAsnThrGlnSerIle 404
Qy 545 GATACATAAGTAATATTACTAGAACTGGGTACTTATTAGATGATCTTATCGTATG 604
Db 405 GluYasnSerLysAlaIleAlaAlaAsnThrArgThrLeuGlnGlnHisSerAlaArg 424
Qy 605 ATGGAACAAAATACACATAATATCAATAAGTTGCTAAAGATTGCAAACTGGTTAGCC 664
Db 425 LeuAspSerGlnArgGlnIleAsnGluAsnHisLysGluMetLysArgAlaAla 444
Qy 665 AACCAATCAGATTGCTATTGTAGTCAACCAATCGTGTAGCAAAACGAGCGTTTCT 724
Db 445 GlnSerAlaAlaLeuThrGlyLeuPheGlnProTyrSerValGlyLysPheAsnAlaThr 464
Qy 725 GCTCGGTAGGAGTTATAGATAAACTGCATTAGCCATTGGTGTGGCTACGCATT 784
Db 465 AlaAlaValGlyGlyTyrSerAspGlnAlaLeuAlaValGlyValGlyTyrArgPhe 484
Qy 785 ACTGATCGCTTACCGTTAAAGCGGTGTAGCTTCAATACCTACATGGCGGCATCTCT 844
Db 485 AsnGluGlnThrAlaAlaLysAlaGlyValAlaPheSer-----AspGlyAspAlaSer 502
Qy 845 TATGGTCTCTGTGGTTATGAATTC 871
Db 503 TrpAsnValGlyValAsnPheGluPhe 511

RESULT 12
Q9LA60 PRELIMINARY; PRT; 392 AA.
AC Q9LA60;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Immunoglobulin-binding protein Eiba.
GN Eiba.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ECOR-9;
RX MEDLINE=20187497; PubMed=10722621;
RA Sandt C.H., Hill C.W.;
RT "Four different genes responsible for nonimmune immunoglobulin-binding
RT activities within a single strain of Escherichia coli.";
RL Infect. Immun. 68:2205-2214 (2000).
```



```

Query Match: 10.60% Indels: 24
DB: 2 Gaps: 3

US-10-030-529A-1 (1-1168) x Q8KH7 (1-362)

QY 479 GATTAAAGACACGAGTATTTAAATTAAGTCTTGATGCACGATATTTCACAAAATAAACAA 538
   |||:||||| ||| |||:|||||
Db 254 AspIleLysAlaAspIleAlaThrAsnLysAlaAsp-----IleAlaLysAsnSerAla 271
QY 539 AATATTGATCTACTAAGTAATAATTTACTAGAACTGGGTACTTTATTAGATGATCTTAT 598
   |||:||||| ||| |||:|||||
Db 272 ArgIleAspSerLeuAspLys----- 278
QY 599 CGTATGATGGACAAATACACATAATCAATTAAGTTGTCTAAAGAAATTCCAAACTGGT 658
   |||:||||| ||| |||:|||||
Db 279 -----AsnValAlaAsnLeuArgLysGluThrArgGlnGly 290
QY 659 TTAGCCCAACCAATCAGCATTTCTATGTTAGTCAACCAATGGTGTAGGCAAAACGAGC 718
   |||:||||| ||| |||:|||||
Db 291 LeuAlaGluGlnAlaAlaLeuSerGlyLeuPheGlnProTyrAsnValGlyArgPheAsn 310
QY 719 GTTCTGCTCGGTAGAGGTTATAGAGATAAACTGCATTAGCCATTGGTGTGCGGTCA 778
   |||:||||| ||| |||:|||||
Db 311 ValThrAlaAlaValGlyGlyTyrLysSerGluSerAlaValAlaIleGlyThrGlyPhe 330
QY 779 CGCATTACTGCTGCTTACCGCTAAAGCGGTCTAGCTTCAATACCTCAATGCGC--- 835
   |||:||||| ||| |||:|||||
Db 331 ArgPheThrGluAsnPheAlaAlaLysAlaGlyValAlaValGlyThrSerSerGlySer 350
QY 836 GGCATGCTTATGGTGTCTGTTGGTTATGAATTC 871
Db 351 SerAlaAlaTyrHisValGlyValAsnTyrGluTrp 362

RESULT 15
Q9JXX7 PRELIMINARY; PRT; 364 AA.
AC
DT 01-OCT-2000 (TrEMBLrel. 15; Created)
DT 01-OCT-2000 (TrEMBLrel. 15; Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24; Last annotation update)
DE Adhesin/invasin, putative.
GN NMB1994.
OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC Neisseriaceae; Neisseria.
OX NCBI_TaxID=491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MC58 / Serogroup B;
RA MEDLINE=20175755; PubMed=10710307;
RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
RA Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
RA Nelson W.C., Gwinn M.L., DeBoy R., Peterson J.D., Hickey E.K.,
RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA Mason T., Ciecko A., Parksey D.S., Blair E., Ciftone H., Clark E.B.,
RA Cotton M.D., Utterback T.R., Khouri H., Qin H., Vamathevan J.,
RA Gill J., Scariato V., Maignani V., Pizzi M., Grandi G., Sun L.,
RA Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;
RT "Complete genome sequence of Neisseria meningitidis serogroup B strain
RT MC58."
RL Science 287:1809-1815 (2000).
DR EMBL; AE002548; AAF42321.1; -.
DR PIR; A81019; A81019.
DR TIGR; NMB1994; -.
DR InterPro; IPR005594; Yada.
DR Pfam; PF03895; Yada; 1.
KW Complete proteome.
SQ SEQUENCE 364 AA; 37972 MW; B9FF591FD25F853F CRC64;

```

Alignment Scores:
Pred. No.: 8.53e-10 Length: 364
Score: 217.00 Matches: 52
Percent Similarity: 53.79% Conservative: 19

Best Local Similarity: 39.39% Mismatches: 37
Query Match: 10.60% Indels: 24
DB: 16 Gaps: 3

US-10-030-529A-1 (1-1168) x Q9JXX7 (1-364)

```

QY 479 GATTAAAGACACGAGTATTTAAATTAAGTCTTGATGCACGATATTTCACAAAATAAACAA 538
   |||:||||| ||| |||:|||||
Db 256 AspIleLysAlaAspIleAlaThrAsnLysAlaAsp-----IleAlaLysAsnSerAla 273
QY 539 AATATTGATCTACTAAGTAATAATTTACTAGAACTGGGTACTTTATTAGATGATCTTAT 598
   |||:||||| ||| |||:|||||
Db 274 ArgIleAspSerLeuAspLys----- 280
QY 599 CGTATGATGGACAAATACACATAATCAATTAAGTTGTCTAAAGAAATTCCAAACTGGT 658
   |||:||||| ||| |||:|||||
Db 281 -----AsnValAlaAsnLeuArgLysGluThrArgGlnGly 292
QY 659 TTAGCCCAACCAATCAGCATTTCTATGTTAGTCAACCAATGGTGTAGGCAAAACGAGC 718
   |||:||||| ||| |||:|||||
Db 293 LeuAlaGluGlnAlaAlaLeuSerGlyLeuPheGlnProTyrAsnValGlyArgPheAsn 312
QY 719 GTTCTGCTCGGTAGAGGTTATAGAGATAAACTGCATTAGCCATTGGTGTGCGGTCA 778
   |||:||||| ||| |||:|||||
Db 313 ValThrAlaAlaValGlyGlyTyrLysSerGluSerAlaValAlaIleGlyThrGlyPhe 332
QY 779 CGCATTACTGCTGCTTACCGCTAAAGCGGTGTAGCTTCAATACCTCAATGCGC--- 835
   |||:||||| ||| |||:|||||
Db 333 ArgPheThrGluAsnPheAlaAlaLysAlaGlyValAlaValGlyThrSerSerGlySer 352
QY 836 GGCATGCTTATGGTGTCTGTTGGTTATGAATTC 871
Db 353 SerAlaAlaTyrHisValGlyValAsnTyrGluTrp 364

```

Search completed: May 13, 2004, 08:28:31
Job time : 73.5 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 13, 2004, 08:14:11 ; Search time 60 Seconds
(without alignments)
1210.245 Million cell updates/sec

Title: US-10-030-529A-2

Perfect score: 1342

Sequence: 1 MKIKCLVAVVGLACSTITM.....VAFTYNGMGSMYGASVGYEF 257

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: Geneseqp_29Jan04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1342	100.0	257	4	AAB31713 Amino aci
2	1342	100.0	257	4	AAB31705 Amino aci
3	1342	100.0	257	4	AAB31709 Amino aci
4	1299.5	96.8	264	4	AAB31706 Amino aci
5	1299.5	96.8	264	4	AAB31707 Amino aci
6	1289	96.1	271	4	AAB31711 Amino aci
7	1283	95.6	273	4	AAB31712 Amino aci
8	1259	93.8	242	4	AAB31708 Amino aci
9	1205	89.8	263	4	AAB31710 Amino aci
10	217	16.2	355	6	ABU07925 Neisseria
11	217	16.2	357	6	ABU07926 Neisseria
12	217	16.2	362	6	ABU07915 Neisseria
13	217	16.2	364	3	AAY75736 Neisseria
14	217	16.2	364	4	AAU27562 Neisseria
15	217	16.2	364	6	ABU07918 Neisseria
16	210	15.6	323	6	ABU07927 Neisseria
17	209	15.6	391	6	ABU07921 Neisseria
18	209	15.6	393	6	ABU07922 Neisseria
19	209	15.6	398	6	ABU07916 Neisseria
20	209	15.6	400	6	ABU07919 Neisseria
21	206	15.4	325	6	ABU07928 Neisseria
22	204	15.2	405	6	ABU07917 Neisseria
23	204	15.2	405	6	ABU07923 Neisseria
24	204	15.2	407	6	ABU07924 Neisseria
25	204	15.2	407	6	ABU07920 Neisseria

26	204	15.2	645	4	AAE10036
27	204	15.2	645	4	AAU27601
28	204	15.2	648	4	AAE10028
29	204	15.2	648	4	AAU27579
30	204	15.2	793	4	AAE10016
31	204	15.2	793	4	AAU27571
32	204	15.2	806	4	AAE10033
33	204	15.2	806	4	AAE10035
34	204	15.2	806	4	AAU27598
35	204	15.2	806	4	AAU27600
36	204	15.2	858	4	AAE10019
37	204	15.2	858	4	AAU27574
38	204	15.2	1444	4	AAE10037
39	204	15.2	1444	4	AAU27602
40	204	15.2	1447	4	AAE10024
41	204	15.2	1447	4	AAU27577
42	193	14.4	610	2	AAW68206
43	193	14.4	624	2	AAW68204
44	193	14.4	889	2	AAW68208
45	190.5	14.2	573	2	AAW68202

ALIGNMENTS

RESULT 1

AAB31713
ID AAB31713 standard; protein; 257 AA.

XX AAB31713;

DT 30-APR-2001 (first entry)

XX Amino acid sequence of the DsrA protein from strain OF406.

DE DsrA protein; outer membrane protein; serum resistance; vaccine;
KW chancroid disease; genital ulcer disease.

XX Haemophilus ducreyi.

XX WO200104138-A1.

PD 18-JAN-2001.

XX 07-JUL-2000; 2000WO-US018834.

XX 09-JUL-1999; 99US-0143257P.

XX (UYNC-) UNIV NORTH CAROLINA.
XX (ELKI/) ELKINS C.

XX Elkins C;

XX WPI; 2001-138311/14.

XX N-PSDB; AAF25270.

XX Novel purified DsrA protein of Haemophilus ducreyi which confers serum resistance to the bacteria. DsrA used to produce vaccines that induce immune response against the bacteria in subject at risk of developing chancroid.

PS Claim 7; Page 58-59; 80pp; English.

XX The present sequence represents a DsrA 30 kDa protein of Haemophilus ducreyi. DsrA is an outer membrane protein that confers serum resistance to the bacteria. DsrA antisense oligonucleotides are useful for detecting a polynucleotide which encodes DsrA in a biological sample. The DsrA polypeptide is used to produce vaccine compositions, which are useful for inducing a protective immune response in a subject at risk of developing Haemophilus ducreyi infection e.g. chancroid, a genital ulcer disease transmitted by sexual contact. DsrA, its catalytic or immunogenic fragments or peptides can be used for screening libraries of compounds in a variety of drug screening techniques. The proteins and peptides may be used as antigens in immunoassays for detection of Haemophilus ducreyi in

CC various tissues and body fluids, e.g. blood, spinal fluid, sputum etc. .
CC The nucleic acids are useful for the preparation of DsrA proteins
XX
SQ Sequence 257 AA;
Query Match 100.0%; Score 1342; DB 4; Length 257;
Best Local Similarity 100.0%; Pred. NO. 6.1e-133;
Matches 257; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MKIKCLVAVVGLACSTTTTMAQQPKFAGVSSLYSEYDYGKWTWNEGDFDIKVPGI 60
DB 1 MKIKCLVAVVGLACSTTTTMAQQPKFAGVSSLYSEYDYGKWTWNEGDFDIKVPGI 60
QY 61 KMKPKEWISKOATYLELQHYMPYTPVLVYAPGVSPSPILLYPMSDDPDLGINRQOLKLN 120
DB 61 KMKPKEWISKOATYLELQHYMPYTPVLVYAPGVSPSPILLYPMSDDPDLGINRQOLKLN 120
QY 121 LYSYFNDLRHDPKLVLDARISKQKQNDITISKYLLELGTLYLDDSYRMEQNTNINKLS 180
DB 121 LYSYFNDLRHDPKLVLDARISKQKQNDITISKYLLELGTLYLDDSYRMEQNTNINKLS 180
QY 181 KELOTGLANQSALSMVQPNVGKTSVSAAGVGYRDKTALAIGVGSRTIDRFTAKAGVAF 240
DB 181 KELOTGLANQSALSMVQPNVGKTSVSAAGVGYRDKTALAIGVGSRTIDRFTAKAGVAF 240
QY 241 NTYNGMSYGASVGYEF 257
DB 241 NTYNGMSYGASVGYEF 257
RESULT 2
AAB31705
ID AAB31705 standard; protein; 257 AA.
XX
AC AAB31705;
XX
DT 30-APR-2001 (first entry)
XX
DE Amino acid sequence of the DsrA locus.
XX
KW DsrA protein; outer membrane protein; serum resistance; vaccine;
KW chancroid disease; genital ulcer disease.
XX
OS Haemophilus ducreyi.
XX
FN WO200104138-A1.
XX
PD 18-JAN-2001.
XX
PF 07-JUL-2000; 2000WO-US018834.
XX
PR 09-JUL-1999; 99US-0143257P.
XX
XX (UYN-) UNIV NORTH CAROLINA.
PA (ELKI/) ELKINS C.
XX
PI Elkins C;
XX
DR WPI; 2001-138311/14.
DR N-PSDB; AAF25262.
XX
XX Novel purified DsrA protein of Haemophilus ducreyi which confers serum
PT resistance to the bacteria used to produce vaccines that induce immune
PT response against the bacteria in subject at risk of developing chancroid.
XX
PS Claim 3; Fig 3; 80pp; English.
XX
CC The present sequence represents a DsrA 30 kDa protein of Haemophilus
CC ducreyi. DsrA is an outer membrane protein that confers serum resistance
CC to the bacteria. DsrA antisense oligonucleotides are useful for detecting
CC a polynucleotide which encodes DsrA in a biological sample. The DsrA
CC polypeptide is used to produce vaccine compositions, which are useful for
CC inducing a protective immune response in a subject at risk of developing

CC Haemophilus ducreyi infection e.g. chancroid, a genital ulcer disease
CC transmitted by sexual contact. DsrA, its catalytic or immunogenic
CC fragments or peptides can be used for screening libraries of compounds in
CC a variety of drug screening techniques. The proteins and peptides may be
CC used as antigens in immunoassays for detection of Haemophilus ducreyi in
CC various tissues and body fluids, e.g. blood, spinal fluid, sputum etc. .
XX The nucleic acids are useful for the preparation of DsrA proteins
XX
SQ Sequence 257 AA;
Query Match 100.0%; Score 1342; DB 4; Length 257;
Best Local Similarity 100.0%; Pred. NO. 6.1e-133;
Matches 257; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MKIKCLVAVVGLACSTTTTMAQQPKFAGVSSLYSEYDYGKWTWNEGDFDIKVPGI 60
DB 1 MKIKCLVAVVGLACSTTTTMAQQPKFAGVSSLYSEYDYGKWTWNEGDFDIKVPGI 60
QY 61 KMKPKEWISKOATYLELQHYMPYTPVLVYAPGVSPSPILLYPMSDDPDLGINRQOLKLN 120
DB 61 KMKPKEWISKOATYLELQHYMPYTPVLVYAPGVSPSPILLYPMSDDPDLGINRQOLKLN 120
QY 121 LYSYFNDLRHDPKLVLDARISKQKQNDITISKYLLELGTLYLDDSYRMEQNTNINKLS 180
DB 121 LYSYFNDLRHDPKLVLDARISKQKQNDITISKYLLELGTLYLDDSYRMEQNTNINKLS 180
QY 181 KELOTGLANQSALSMVQPNVGKTSVSAAGVGYRDKTALAIGVGSRTIDRFTAKAGVAF 240
DB 181 KELOTGLANQSALSMVQPNVGKTSVSAAGVGYRDKTALAIGVGSRTIDRFTAKAGVAF 240
QY 241 NTYNGMSYGASVGYEF 257
DB 241 NTYNGMSYGASVGYEF 257
RESULT 3
AAB31709
ID AAB31709 standard; protein; 257 AA.
XX
AC AAB31709;
XX
DT 30-APR-2001 (first entry)
XX
DE Amino acid sequence of the DsrA protein from strain CIP542 (CDC) .
XX
KW DsrA protein; outer membrane protein; serum resistance; vaccine;
KW chancroid disease; genital ulcer disease.
XX
OS Haemophilus ducreyi.
XX
FN WO200104138-A1.
XX
PD 18-JAN-2001.
XX
PF 07-JUL-2000; 2000WO-US018834.
XX
PR 09-JUL-1999; 99US-0143257P.
XX
XX (UYN-) UNIV NORTH CAROLINA.
PA (ELKI/) ELKINS C.
XX
PI Elkins C;
XX
DR WPI; 2001-138311/14.
DR N-PSDB; AAF25266.
XX
XX Novel purified DsrA protein of Haemophilus ducreyi which confers serum
PT resistance to the bacteria used to produce vaccines that induce immune
PT response against the bacteria in subject at risk of developing chancroid.
XX
PS Claim 7; Page 55; 80pp; English.
XX
CC The present sequence represents a DsrA 30 kDa protein of Haemophilus

CC ducreyi. DsrA is an outer membrane protein that confers serum resistance
 CC to the bacteria. DsrA antisense oligonucleotides are useful for detecting
 CC a polynucleotide which encodes DsrA in a biological sample. The DsrA
 CC polypeptide is used to produce vaccine compositions, which are useful for
 CC inducing a protective immune response in a subject at risk of developing
 CC Haemophilus ducreyi infection e.g. chancroid, a genital ulcer disease
 CC transmitted by sexual contact. DsrA, its catalytic or immunogenic
 CC fragments or peptides can be used for screening libraries of compounds in
 CC a variety of drug screening techniques. The proteins and peptides may be
 CC used as antigens in immunoassays for detection of Haemophilus ducreyi in
 CC various tissues and body fluids, e.g. blood, spinal fluid, sputum etc. .
 CC The nucleic acids are useful for the preparation of DsrA proteins
 XX
 SQ Sequence 257 AA;

Query Match 100.0%; Score 1342; DB 4; Length 257;
 Best Local Similarity 100.0%; Pred. No. 6.1e-133;
 Matches 257; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MKIKCLVAVVGLACSTTTTWAQQPPKPGAGVSSLYSYEDYDGKWTWSNEGDFDIKVPGI 60
 DB 1 MKIKCLVAVVGLACSTTTTWAQQPPKPGAGVSSLYSYEDYDGKWTWSNEGDFDIKVPGI 60
 QY 61 KMKPEWISKQATYLELQHYMPYTPVLVYAPGVSPSPILLYPMSDDPQLGINRQQLKLN 120
 DB 61 KMKPEWISKQATYLELQHYMPYTPVLVYAPGVSPSPILLYPMSDDPQLGINRQQLKLN 120
 QY 121 LYSYFNDLRHDFKLVLDARISKKNQNDITISKYLLLELGTLYDSDSYRMMEQNTNINKLS 180
 DB 121 LYSYFNDLRHDFKLVLDARISKKNQNDITISKYLLLELGTLYDSDSYRMMEQNTNINKLS 180
 QY 181 KELQTGLANQSALSMVQPNVGKTSVSAAGVGYRDKTALAIGVGSRTDRFTAKAGVAF 240
 DB 181 KELQTGLANQSALSMVQPNVGKTSVSAAGVGYRDKTALAIGVGSRTDRFTAKAGVAF 240
 QY 241 NTYNGMSYGASVGVEF 257
 DB 241 NTYNGMSYGASVGVEF 257

RESULT 4
 AAB31706
 ID AAB31706 standard; protein; 264 AA.

XX AAB31706;
 XX
 XX 30-APR-2001 (first entry)
 XX
 XX Amino acid sequence of the DsrA protein from strain CIPA75.

XX DsrA protein; outer membrane protein; serum resistance; vaccine;
 XX chancroid disease; genital ulcer disease.

XX Haemophilus ducreyi.

XX WO200104138-A1.

XX 18-JAN-2001.

XX 07-JUL-2000; 2000WO-US018834.

XX 09-JUL-1999; 99US-0143257P.

XX (UYNC-) UNIV NORTH CAROLINA.

XX (ELKI/) ELKINS C.

XX Elkins C;

XX WPI; 2001-138311/14.

XX N-PSDB; AAF25263.

XX Novel purified DsrA protein of Haemophilus ducreyi which confers serum
 PT resistance to the bacteria used to produce vaccines that induce immune

PT response against the bacteria in subject at risk of developing chancroid.
 XX
 PS Claim 7; Page 53; 80pp; English.

XX The present sequence represents a DsrA 30 kDa protein of Haemophilus
 CC ducreyi. DsrA is an outer membrane protein that confers serum resistance
 CC to the bacteria. DsrA antisense oligonucleotides are useful for detecting
 CC a polynucleotide which encodes DsrA in a biological sample. The DsrA
 CC polypeptide is used to produce vaccine compositions, which are useful for
 CC inducing a protective immune response in a subject at risk of developing
 CC Haemophilus ducreyi infection e.g. chancroid, a genital ulcer disease
 CC transmitted by sexual contact. DsrA, its catalytic or immunogenic
 CC fragments or peptides can be used for screening libraries of compounds in
 CC a variety of drug screening techniques. The proteins and peptides may be
 CC used as antigens in immunoassays for detection of Haemophilus ducreyi in
 CC various tissues and body fluids, e.g. blood, spinal fluid, sputum etc. .
 CC The nucleic acids are useful for the preparation of DsrA proteins
 XX
 SQ Sequence 264 AA;

Query Match 96.8%; Score 1299.5; DB 4; Length 264;
 Best Local Similarity 95.8%; Pred. No. 2e-128;
 Matches 253; Conservative 0; Mismatches 4; Indels 7; Gaps 1;
 QY 1 MKIKCLVAVVGLACSTTTTWAQQPPKPGAGVSSLYSYEDYDGKWTWSNEGDFDIKVPGI 60
 DB 1 MKIKCLVAVVGLACSTTTTWAQQPPKPGAGVSSLYSYEDYDGKWTWSNEGDFDIKVPGI 60
 QY 61 KMKPEWISKQATYLELQHYMPYTPVLVYAPGVSPSPILLYPMSDDPQLGINRQQLKLN 120
 DB 61 KMKPEWISKQATYLELQHYMPYTPVLVYAPGVSPSPILLYPMSDDPQLGINRQQLKLN 120
 QY 121 LYSYFNDLRHDFKLVLDARISKKNQNDITISKYLLLELGTLYDSDSYRMMEQNTNINKNT 173
 DB 121 LYSYFNDLRHDFKLVLDARISKKNQNDITISKYLLLELGTLYDSDSYRMMEQNTNINKNT 180
 QY 174 HNINKLSKELQGTGLANQSALSMVQPNVGKTSVSAAGVGYRDKTALAIGVGSRTDRFT 233
 DB 181 HNINKLSKELQGTGLANQSALSMVQPNVGKTSVSAAGVGYRDKTALAIGVGSRTDRFT 240
 QY 234 AKAGVAFNTYNGMSYGASVGVEF 257
 DB 241 AKAGVAFNTYNGMSYGASVGVEF 264

RESULT 5
 AAB31707
 ID AAB31707 standard; protein; 264 AA.

XX AAB31707;
 XX
 XX 30-APR-2001 (first entry)
 XX
 XX Amino acid sequence of the DsrA protein from strain CIPA77.

XX DsrA protein; outer membrane protein; serum resistance; vaccine;
 XX chancroid disease; genital ulcer disease.
 XX Haemophilus ducreyi.

XX WO200104138-A1.

XX 18-JAN-2001.

XX 07-JUL-2000; 2000WO-US018834.

XX 09-JUL-1999; 99US-0143257P.

XX (UYNC-) UNIV NORTH CAROLINA.

XX (ELKI/) ELKINS C.

XX Elkins C;

XX

```

DR WPI; 2001-138311/14.
DR N-PSDB; AAF25264.
XX
PT Novel purified DsrA protein of Haemophilus ducreyi which confers serum
PT resistance to the bacteria used to produce vaccines that induce immune
PT response against the bacteria in subject at risk of developing chancroid.
XX
PS Claim 7; Page 54; 80pp; English.
XX
XX The present sequence represents a DsrA 30 kDa protein of Haemophilus
CC ducreyi. DsrA is an outer membrane protein that confers serum resistance
CC to the bacteria. DsrA antisense oligonucleotides are useful for detecting
CC a polynucleotide which encodes DsrA in a biological sample. The DsrA
CC polypeptide is used to produce vaccine compositions, which are useful for
CC inducing a protective immune response in a subject at risk of developing
CC Haemophilus ducreyi infection e.g. chancroid, a genital ulcer disease
CC transmitted by sexual contact. DsrA, its catalytic or immunogenic
CC fragments or peptides can be used for screening libraries of compounds in
CC a variety of drug screening techniques. The proteins and peptides may be
CC used as antigens in immunoassays for detection of Haemophilus ducreyi in
CC various tissues and body fluids, e.g. blood, spinal fluid, sputum etc. .
CC The nucleic acids are useful for the preparation of DsrA proteins
XX
SQ Sequence 264 AA;
Query Match 96.8%; Score 1299.5; DB 4; Length 264;
Best Local Similarity 95.8%; Pred. No. 2e-128;
Matches 253; Conservative 0; Mismatches 4; Indels 7; Gaps 1;
QY 1 MKIKLVAVVGLACSTTTTMAQQPKFAGVSSLSYSEYDYGKGTWNSGEGFDIKVPGI 60
Db 1 MKIKLVAVVGLACSTTTTMAQQPKFAGVSSLSYSEYDYGKGTWNSGEGFDIKVPGI 60
QY 61 KMKPKEWISKQATYLELOHYMPTVLTVPVLYAPGVSPSPILLYPMSDDPOLGINRQOLKLN 120
Db 61 KMKPKEWISKQATYLELOHYMPTVLTVPVLYAPGVSPSPILLYPMSDDPOLGINRQOLKLN 120
QY 121 LYSYFNDLRHDFKLVLDARISKKNQNDITISKYLLELGTLYLDDSYRMEQ-----NT 173
Db 121 LYSYFNDLRHDFKLVLDARISKKNQNDITISKYLLELGTLYLDDSYRMEQNTNINKNT 180
QY 174 HNINKLSKELOTGLANQSALSMVQPNVGKTSVSAAVGGYRDKTALAIGVGSRTIDRFT 233
Db 181 HNINKLSKELOTGLANQSALSMVQPNVGKTSVSAAVGGYRDKTALAIGVGSRTIDRFT 240
QY 234 AKAGVAFNTYNGGMSYGASVGYEF 257
Db 241 AKAGVAFNTYNGGMSYGASVGYEF 264
RESULT 6
AAB31711
ID AAB31711 standard; protein; 271 AA.
XX
XX AAB31711;
XX
XX 30-APR-2001 (first entry)
XX
DE Amino acid sequence of the DsrA protein from strain V-1157.
XX
XX DsrA protein; outer membrane protein; serum resistance; vaccine;
XX chancroid disease; genital ulcer disease.
XX
OS Haemophilus ducreyi.
XX
XX WO200104138-A1.
XX
XX 18-JAN-2001.
XX
XX 07-JUL-2000; 2000WO-US018834.
XX
XX 09-JUL-1999; 99US-0143257P.
XX
XX

```

```

PA (UYNC-) UNIV NORTH CAROLINA.
PA (ELKI/) ELKINS C.
XX
PI Elkins C;
XX
XX WPI; 2001-138311/14.
DR N-PSDB; AAF25268.
XX
XX Novel purified DsrA protein of Haemophilus ducreyi which confers serum
PT resistance to the bacteria used to produce vaccines that induce immune
PT response against the bacteria in subject at risk of developing chancroid.
XX
PS Claim 7; Page 57; 80pp; English.
XX
XX The present sequence represents a DsrA 30 kDa protein of Haemophilus
CC ducreyi. DsrA is an outer membrane protein that confers serum resistance
CC to the bacteria. DsrA antisense oligonucleotides are useful for detecting
CC a polynucleotide which encodes DsrA in a biological sample. The DsrA
CC polypeptide is used to produce vaccine compositions, which are useful for
CC inducing a protective immune response in a subject at risk of developing
CC Haemophilus ducreyi infection e.g. chancroid, a genital ulcer disease
CC transmitted by sexual contact. DsrA, its catalytic or immunogenic
CC fragments or peptides can be used for screening libraries of compounds in
CC a variety of drug screening techniques. The proteins and peptides may be
CC used as antigens in immunoassays for detection of Haemophilus ducreyi in
CC various tissues and body fluids, e.g. blood, spinal fluid, sputum etc. .
CC The nucleic acids are useful for the preparation of DsrA proteins
XX
SQ Sequence 271 AA;
Query Match 96.1%; Score 1289; DB 4; Length 271;
Best Local Similarity 93.0%; Pred. No. 2.6e-127;
Matches 252; Conservative 0; Mismatches 5; Indels 14; Gaps 1;
QY 1 MKIKLVAVVGLACSTTTTMAQQPKFAGVSSLSYSEYDYGKGTWNSGEGFDIKVPGI 60
Db 1 MKIKLVAVVGLACSTTTTMAQQPKFAGVSSLSYSEYDYGKGTWNSGEGFDIKVPGI 60
QY 61 KMKPKEWISKQATYLELOHYMPTVLTVPVLYAPGVSPSPILLYPMSDDPOLGINRQOLKLN 120
Db 61 KMKPKEWISKQATYLELOHYMPTVLTVPVLYAPGVSPSPILLYPMSDDPOLGINRQOLKLN 120
QY 121 LYSYFNDLRHDFKLVLDARISKKNQNDITISKYLLELGTLYLDDSYRMEQ----- 171
Db 121 LYSYFNDLRHDFKLVLDARISKKNQNDITISKYLLELGTLYLDDSYRMEQNTNINKNT 180
QY 172 -----NTHNINKLSKELOTGLANQSALSMVQPNVGKTSVSAAVGGYRDKTALAIGVGS 226
Db 181 HNINKNTHNINKLSKELOTGLANQSALSMVQPNVGKTSVSAAVGGYRDKTALAIGVGS 240
QY 227 RITDRFTAKAGVAFNTYNGGMSYGASVGYEF 257
Db 241 RITDRFTAKAGVAFNTYNGGMSYGASVGYEF 271
RESULT 7
AAB31712
ID AAB31712 standard; protein; 273 AA.
XX
XX AAB31712;
XX
XX 30-APR-2001 (first entry)
XX
DE Amino acid sequence of the DsrA protein from strain M90-02.
XX
XX DsrA protein; outer membrane protein; serum resistance; vaccine;
XX chancroid disease; genital ulcer disease.
XX
OS Haemophilus ducreyi.
XX
XX WO200104138-A1.
XX
XX 18-JAN-2001.
XX

```

```

XX PF 07-JUL-2000; 2000WO-US018834.
XX PR 09-JUL-1999; 99US-0143257P.
XX XX (UYNK-) UNIV NORTH CAROLINA.
XX PA (ELKI/) ELKINS C.
XX XX
XX PI Elkins C;
XX XX
XX DR WPI; 2001-138311/14.
XX DR N-PSDB; AAF25269.
XX XX
XX PT Novel purified DsrA protein of Haemophilus ducreyi which confers serum
XX PT resistance to the bacteria used to produce vaccines that induce immune
XX PT response against the bacteria in subject at risk of developing chancroid.
XX PS
XX PS Claim 7; Page 58; 80pp; English.
XX CC
XX CC The present sequence represents a DsrA 30 kDa protein of Haemophilus
XX CC ducreyi. DsrA is an outer membrane protein that confers serum resistance
XX CC to the bacteria. DsrA antisense oligonucleotides are useful for detecting
XX CC a polynucleotide which encodes DsrA in a biological sample. The DsrA
XX CC polypeptide is used to produce vaccine compositions, which are useful for
XX CC inducing a protective immune response in a subject at risk of developing
XX CC Haemophilus ducreyi infection e.g. chancroid, a genital ulcer disease
XX CC transmitted by sexual contact. DsrA, its catalytic or immunogenic
XX CC fragments or peptides can be used for screening libraries of compounds in
XX CC a variety of drug screening techniques. The proteins and peptides may be
XX CC used as antigens in immunoassays for detection of Haemophilus ducreyi in
XX CC various tissues and body fluids, e.g. blood, spinal fluid, sputum etc.
XX CC The nucleic acids are useful for the preparation of DsrA proteins
XX XX
XX SQ Sequence 273 AA;
XX
Query Match 95.6%; Score 1283; DB 4; Length 273;
Best Local Similarity 92.7%; Pred. No. 1.1e-126;
Matches 253; Conservative 0; Mismatches 4; Indels 16; Gaps 2;
QY 1 MKIKLVAVVGLACSTTTTMAQQPPKPGVSSLYSYDYDGKGTWTSNEGPFDIKVPGI 60
DB 1 MKIKLVAVVGLACSTTTTMAQQPPKPGVSSLYSYDYDGKGTWTSNEGPFDIKVPGI 60
QY 61 KMKPKWISKQATYLELQHYMPYTPVLVTYAGVSPS--PILLYPMSDDPDQGINRQOLK 118
DB 61 KMKPKWISKQATYLELQHYMPYTPVLVTYAGVSPS--PILLYPMSDDPDQGINRQOLK 120
QY 119 LNLVSYFNDLRHDFKLVLDARISKKNQIDTISKYLLLELGTYLDDSVRMWQ----- 171
DB 121 LNLVSYFNDLRHDFKLVLDARISKKNQIDTISKYLLLELGTYLDDSVRMWQTHNINK 180
QY 172 -----NTHNINKLSKELOTGLANQSALSMVQPNGVGKTSVSAAGVGYRDKTALAIGV 224
DB 181 NTHNINKNTHNINKLSKELOTGLANQSALSMVQPNGVGKTSVSAAGVGYRDKTALAIGV 240
QY 225 GSRITDRFTAKAGVAFNTYNGGMSYGASVGYEF 257
DB 241 GSRITDRFTAKAGVAFNTYNGGMSYGASVGYEF 273
XX
RESULT 8
AAB31708
ID AAB31708 standard; protein; 242 AA.
XX AC
XX AC AAB31708;
XX XX
XX DT 30-APR-2001 (first entry)
XX DE
XX DE Amino acid sequence of the DsrA protein from strain CIP542 (Can).
XX XX
XX KW DsrA protein; outer membrane protein; serum resistance; vaccine;
XX KW chancroid disease; genital ulcer disease.

```

```

OS Haemophilus ducreyi.
XX WO200104138-A1.
XX PD 18-JAN-2001.
XX PF 07-JUL-2000; 2000WO-US018834.
XX PR 09-JUL-1999; 99US-0143257P.
XX XX (UYNK-) UNIV NORTH CAROLINA.
XX PA (ELKI/) ELKINS C.
XX XX
XX PI Elkins C;
XX XX
XX DR WPI; 2001-138311/14.
XX DR N-PSDB; AAF25265.
XX XX
XX PT Novel purified DsrA protein of Haemophilus ducreyi which confers serum
XX PT resistance to the bacteria used to produce vaccines that induce immune
XX PT response against the bacteria in subject at risk of developing chancroid.
XX PS
XX PS Claim 7; Page 55; 80pp; English.
XX CC
XX CC The present sequence represents a DsrA 30 kDa protein of Haemophilus
XX CC ducreyi. DsrA is an outer membrane protein that confers serum resistance
XX CC to the bacteria. DsrA antisense oligonucleotides are useful for detecting
XX CC a polynucleotide which encodes DsrA in a biological sample. The DsrA
XX CC polypeptide is used to produce vaccine compositions, which are useful for
XX CC inducing a protective immune response in a subject at risk of developing
XX CC Haemophilus ducreyi infection e.g. chancroid, a genital ulcer disease
XX CC transmitted by sexual contact. DsrA, its catalytic or immunogenic
XX CC fragments or peptides can be used for screening libraries of compounds in
XX CC a variety of drug screening techniques. The proteins and peptides may be
XX CC used as antigens in immunoassays for detection of Haemophilus ducreyi in
XX CC various tissues and body fluids, e.g. blood, spinal fluid, sputum etc.
XX CC The nucleic acids are useful for the preparation of DsrA proteins
XX XX
XX SQ Sequence 242 AA;
XX
Query Match 93.8%; Score 1259; DB 4; Length 242;
Best Local Similarity 100.0%; Pred. No. 3.3e-124;
Matches 242; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MKIKLVAVVGLACSTTTTMAQQPPKPGVSSLYSYDYDGKGTWTSNEGPFDIKVPGI 60
DB 1 MKIKLVAVVGLACSTTTTMAQQPPKPGVSSLYSYDYDGKGTWTSNEGPFDIKVPGI 60
QY 61 KMKPKWISKQATYLELQHYMPYTPVLVTYAGVSPSPILLYPMSDDPDQGINRQOLKLN 120
DB 61 KMKPKWISKQATYLELQHYMPYTPVLVTYAGVSPSPILLYPMSDDPDQGINRQOLKLN 120
QY 121 LYSVFNDLRHDFKLVLDARISKKNQIDTISKYLLLELGTYLDDSVRMWQTHNINKLS 180
DB 121 LYSVFNDLRHDFKLVLDARISKKNQIDTISKYLLLELGTYLDDSVRMWQTHNINKLS 180
QY 181 KELOTGLANQSALSMVQPNGVGKTSVSAAGVGYRDKTALAIGVSGSITDRFTAKAGVAF 240
DB 181 KELOTGLANQSALSMVQPNGVGKTSVSAAGVGYRDKTALAIGVSGSITDRFTAKAGVAF 240
QY 241 NT 242
DB 241 NT 242
XX
RESULT 9
AAB31710
ID AAB31710 standard; protein; 263 AA.
XX AC
XX AC AAB31710;
XX XX
XX DT 30-APR-2001 (first entry)
XX XX

```

DE Amino acid sequence of the DsrA protein from strain CHIA.
XX DsrA protein; outer membrane protein; serum resistance; vaccine;
KW chancroid disease; genital ulcer disease.
XX
OS Haemophilus ducreyi.
XX
XX WO200104138-A1.
XX
XX 18-JAN-2001.
FD
XX 07-JUL-2000; 2000WO-US018834.
PF
XX 09-JUL-1999; 99US-0143257P.
XX
XX (UNNC-) UNIV NORTH CAROLINA.
PA
XX (ELKI/) ELKINS C.
XX
XX Elkins C;
PI
XX
XX WPI; 2001-138311/14.
DR
XX N-PSDB; AAF25267.
DR
XX
XX Novel purified DsrA protein of Haemophilus ducreyi which confers serum
PT resistance to the bacteria used to produce vaccines that induce immune
PT response against the bacteria in subject at risk of developing chancroid.
XX
XX Claim 7; Page 56; 80pp; English.
XX
XX The present sequence represents a DsrA 30 kDa protein of Haemophilus
CC ducreyi. DsrA is an outer membrane protein that confers serum resistance
CC to the bacteria. DsrA antisense oligonucleotides are useful for detecting
CC a polynucleotide which encodes DsrA in a biological sample. The DsrA
CC polypeptide is used to produce vaccine compositions, which are useful for
CC inducing a protective immune response in a subject at risk of developing
CC Haemophilus ducreyi infection e.g. chancroid, a genital ulcer disease
CC transmitted by sexual contact. DsrA, its catalytic or immunogenic
CC fragments or peptides can be used for screening libraries of compounds in
CC a variety of drug screening techniques. The proteins and peptides may be
CC used as antigens in immunoassays for detection of Haemophilus ducreyi in
CC various tissues and body fluids, e.g. blood, spinal fluid, sputum etc.
CC The nucleic acids are useful for the preparation of DsrA proteins
XX
XX Sequence 263 AA;
SQ
Query Match 89.8%; Score 1205; DB 4; Length 263;
Best Local Similarity 89.0%; Pred. No. 1.9e-118;
Matches 235; Conservative 8; Mismatches 13; Indels 8; Gaps 2;
QY 1 MKIKLVAVVGLACSTITTTAAQPPKAGVSSLSYVEYDYGKGTWWSNEGGFDIKVPGI 60
Db 1 MKIKLVAVVGLACSTITTTAAQPPKAGVSSLSYVEYDYGKGTWWSNEGGFDIKAPGI 60
QY 61 KMKPKWISQATYLELQHYMPYTPVLVTVYAPGVSPSPILLYPMDPDQLGINRQQLKLN 120
Db 61 KMKPKWISQATYLELQHYMPYTPVLVTVY-A-SAEPTVLLYPMDPDQLGINRQQLKLN 119
QY 121 LYSYFNDLRHDFKLVLDARISKKNQIDITISKYLLLEGLTYLDDSYRMEQ-----NT 173
Db 120 LYSYFNDLRHGFKLVLDARISKKNQIDITISEYLLKLGTYLDDSYRMEQNTNINMKT 179
QY 174 HNINKLSKELOTGLANOSLSMLVQPNVGKTSVSAAGVGYRDKTALAIGVGSRTIDRFT 233
Db 180 HNINKLSKELOTGLANOSLSMLVQPNVGKTSVSAAGVGYRDKTALAIGVGSRTIDRFT 239
QY 234 AKAGVAFNTYNGMSYGSASVGVEF 257
Db 240 AKAGVAFNTYNGMSYGSASVGVEF 263
RESULT 10
ABU07925
ID ABU07925 standard; protein; 355 AA.

XX AC ABU07925;
XX 23-MAY-2003 (first entry)
DT
XX
XX Neisserial adhesin A (NadA) allele 1/2 chimera (strain 95330).
DE
XX
XX Neisserial adhesin A; NadA; antibacterial; immunostimulant; vaccine;
KW neisserial infection; meningitis; bacterial meningitis; bacteraemia;
KW systemic immunity; mucosal immunity; allele; strain 95330; chimeric.
XX
XX Neisseria meningitidis.
OS Synthetic.
XX
XX WO2003010194-A2.
PN
XX 06-FEB-2003.
PD
XX
XX 26-JUL-2002; 2002WO-IB003396.
PF
XX
XX 27-JUL-2001; 2001GB-00018401.
PR
XX 06-SEP-2001; 2001GB-00021591.
PR
XX 14-MAY-2002; 2002GB-00011025.
XX
XX (CHIR-) CHIRON SPA.
PA
XX
XX Arico M, Comanducci M;
PI
XX
XX WPI; 2003-248057/24.
DR
XX
XX New Neisserial adhesin A protein and nucleic acids, useful for preventing
PT or treating meningitis, particularly bacterial meningitis, and
PT bacteraemia, and for eliciting an systemic and/or mucosal immunity.
XX
XX Claim 1; Page 76; 79pp; English.
XX
XX The invention describes a Neisserial adhesin (NadA) comprising a 362,
CC 398, 405, 364, 400, 407, 391, 393, 405, 107, 355, 357, 323, or 319
CC residue amino acid sequence given in the specification, or an amino acid
CC sequence having at least 50 % identity to the amino acid sequences, or a
CC fragment of them. The NadA protein, or nucleic acid encoding NadA protein
CC is useful in the manufacture of a medication for preventing Neisserial
CC infection in a mammal, such as an infection of Neisseria meningitidis
CC from hypervirulent lineages ET-5, EV-37 and cluster A4. The NadA protein
CC is useful for preventing or treating diseases, specifically meningitis
CC (particularly bacterial meningitis) and bacteraemia, and for eliciting an
CC systemic and/or mucosal immunity. This is the amino acid sequence of a
CC neisserial adhesin A (NadA) allele 1/2 chimera (strain 95330)
XX
XX Sequence 355 AA;
SQ
Query Match 16.2%; Score 217; DB 6; Length 355;
Best Local Similarity 39.4%; Pred. No. 8.6e-14;
Matches 52; Conservative 19; Mismatches 37; Indels 24; Gaps 3;
QY 127 DLRHDFKLVLDARISKKNQIDITISKYLLLEGLTYLDDSYRMEQNTNINKLSKELOTG 186
Db 247 DIKADIATNKAD--IAKNSARIDSLDK-----NVALRKETRQG 283
QY 187 LANQSALSMVQPNVGKTSVSAAGVGYRDKTALAIGVGSRTIDRFTAKAGVAFNTYNG- 245
Db 284 LAEQALSGLFQPNVNGRFTVTAAGVGYKSESVAIGTGFRPTENFAAKAGVAVGTSGS 343
QY 246 GMSYGSASVGVEF 257
Db 344 SAAYHGVNVEW 355
RESULT 11
ABU07926
ID ABU07926 standard; protein; 357 AA.
XX
XX AC ABU07926;

KW Neisserial adhesin A; Nada; antibacterial; immunostimulant; vaccine;
KW neisserial infection; meningitis; bacterial meningitis; bacteraemia;
KW systemic immunity; mucosal immunity; allele.
OS Neisseria meningitidis.
XX WO2003010194-A2.
XX 06-FEB-2003.
XX 26-JUL-2002; 2002WO-IB003396.
XX 27-JUL-2001; 2001GB-00018401.
PR 06-SEP-2001; 2001GB-00021591.
PR 14-MAY-2002; 2002GB-00011025.
XX (CHIR-) CHIRON SPA.
XX Arico M, Comanducci M;
PI WPI; 2003-248057/24.
DR
XX
XX
PT New Neisserial adhesin A protein and nucleic acids, useful for preventing
PT or treating meningitis, particularly bacterial meningitis, and
PT bacteremia, and for eliciting an systemic and/or mucosal immunity.
XX
XX
PS Claim 1; Page 75; 79pp; English.
XX
XX The invention describes a Neisserial adhesin (Nada) comprising a 362,
CC 398, 405, 364, 400, 407, 391, 393, 405, 107, 355, 357, 323, or 319
CC residue amino acid sequence given in the specification, or an amino acid
CC sequence having at least 50 % identity to the amino acid sequences, or a
CC fragment of them. The Nada protein, or nucleic acid encoding Nada protein
CC is useful in the manufacture of a medicament for preventing Neisserial
CC infection in a mammal, such as an infection of Neisseria meningitidis
CC from hypervirulent lineages ET-5, EY-37 and cluster A4. The Nada protein
CC is useful for preventing or treating diseases, specifically meningitis
CC (particularly bacterial meningitis) and bacteraemia, and for eliciting an
CC systemic and/or mucosal immunity. This is the amino acid sequence of
CC neisserial adhesin A (Nada) allele 1 (first ATG start)
XX
SQ Sequence 364 AA;

Query Match 16.2%; Score 217; DB 6; Length 364;
Best Local Similarity 39.4%; Pred. No. 9e-14;
Matches 52; Conservative 19; Mismatches 37; Indels 24; Gaps 3;
Qy 127 DLRHDFKLKVLDAKISKNQKIDITISKYLLLELGYLDDSYRMMEQNTHNINKLSKEIQTG 186
Db 256 DIRADIATNKAD--IAKNSARIDSLDK-----NVAQLRKETROG 292
Qy 187 LANQSALSMVLQPNVGKTSVSAVGVGRDKTALAIGVGSRTIDRFTAKAGVAFNTYNG- 245
Db 293 LAEQALSGFLFQPNVGVGRFNTAAVGVGKSESVAIGTGRFTENFAAKAGVAVGTSSGS 352
Qy 246 GMSYGASVGYEF 257
Db 353 SAAYHVGVNVEW 364

Search completed: May 13, 2004, 08:15:26
Job time : 61 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 13, 2004, 08:14:12 ; Search time 22 Seconds
(without alignments)
603.085 Million cell updates/sec

Title: US-10-030-529a-2

Perfect score: 1342

Sequence: 1 MKIKLVAVVGLACSTITM.....VAFTYNGMSYGASGVYEF 257

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA:*
1: /cgn2_6/ptodata/2/iaa/5A COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	193	14.4	610	4	US-09-336-447A-11
2	193	14.4	624	4	US-09-336-447A-7
3	193	14.4	889	4	US-09-336-447A-15
4	192	14.3	867	4	US-09-540-236-2676
5	190.5	14.2	573	4	US-09-336-447A-3
6	125.5	9.4	568	4	US-09-543-681A-6966
7	115	8.6	2042	4	US-09-077-098A-6
8	114	8.5	2039	4	US-09-077-098A-7
9	110	8.2	873	4	US-09-336-447A-13
10	108	8.0	892	4	US-09-336-447A-5
11	107	8.0	831	4	US-09-336-447A-1
12	107	8.0	878	4	US-09-540-236-3401
13	107	8.0	941	4	US-09-336-447A-9
14	104	7.7	1002	4	US-09-268-347-24
15	104	7.7	1004	4	US-09-268-347-30
16	103.5	7.7	616	4	US-09-268-347-38
17	103	7.7	1104	4	US-09-268-347-28
18	103	7.7	1104	4	US-09-268-347-15
19	102	7.6	679	3	US-08-913-942-15
20	102	7.6	679	4	US-09-268-347-26
21	102	7.6	1094	4	US-09-268-347-32
22	100.5	7.5	1098	1	US-08-409-995-2
23	100.5	7.5	1098	3	US-08-685-467-2
24	100.5	7.5	1098	3	US-09-377-155-32
25	100.5	7.5	1098	3	US-08-913-942-2
26	100.5	7.5	1098	4	US-09-669-974-32
27	100.5	7.5	1098	4	US-09-268-347-44

```

28 100.5 7.5 1098 4 US-09-797-862-32 Sequence 32, Appl
29 100.5 7.5 2353 3 US-09-377-355-33 Sequence 33, Appl
30 100.5 7.5 2353 3 US-08-913-942-4 Sequence 4, Appl
31 100.5 7.5 2353 4 US-09-669-974-33 Sequence 33, Appl
32 100.5 7.5 2353 4 US-09-797-862-33 Sequence 33, Appl
33 100.5 7.5 2411 4 US-09-268-347-36 Sequence 36, Appl
34 99 7.4 1690 4 US-09-595-684B-39 Sequence 39, Appl
35 99 7.4 2354 4 US-09-268-347-47 Sequence 47, Appl
36 93 6.9 1117 2 US-08-843-530B-33 Sequence 33, Appl
37 92 6.9 447 4 US-09-543-681A-6231 Sequence 6231, Ap
38 90 6.7 512 4 US-09-489-039A-12836 Sequence 12836, A
39 88 6.6 776 4 US-10-181-660-5 Sequence 5, Appl
40 88 6.6 776 4 US-10-181-660-13 Sequence 13, Appl
41 88 6.6 797 4 US-09-994-192-4 Sequence 4, Appl
42 88 6.6 797 4 US-10-181-660-3 Sequence 3, Appl
43 86.5 6.6 797 4 US-10-181-660-11 Sequence 11, Appl
44 1226 1 US-08-280-443-2 Sequence 2, Appl
45 86.5 6.4 1226 1 US-08-457-459-2 Sequence 2, Appl

```

ALIGNMENTS

```

RESULT 1
US-09-336-447A-11
; Sequence 11, Application US/09336447A
; Patent No. 6310190
; GENERAL INFORMATION:
; APPLICANT: HANSEN, ERIC J.
; APPLICANT: AEBI, CHRISTOPH
; APPLICANT: COPE, LESLIE D.
; APPLICANT: MACIVER, ISOBEL
; APPLICANT: FISKE, MICHAEL J.
; APPLICANT: FREDENBURG, ROSS A.
; TITLE OF INVENTION: USPA1 AND USPA2 ANTIGENS OF MORAXELLA CATARRHALIS
; FILE REFERENCE: AMCY.024
; CURRENT APPLICATION NUMBER: US/09/336,447A
; CURRENT FILING DATE: 1999-06-21
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 11
; LENGTH: 610
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-09-336-447A-11

```

```

Query Match 14.4%; Score 193; DB 4; Length 610;
Best Local Similarity 37.9%; Pred. No. 4.2e-12;
Matches 53; Conservative 18; Mismatches 61; Indels 8; Gaps 3;

Qy 126 NDLRHDFKLVLDARISKNQNTIDTISKYLLLEGTYLDD-SYRMEQNTH-----NINK 178
Db 471 NKASADTKFAATADAITKNGNAITKNAKSITDLGTVKDFGRVTALDTKVNAFDGRITA 530

Qy 179 LSKELQTLANQSNLSMLVQPNVGKTSVSNANQGYRDKTALAIGVCSRTIDRTAKGV 238
Db 531 LSKVNGMAQAALSGLFQPSYGVKNATAALGGYSGSAVAIGAGYRVPNPNLAFKAGA 590

Qy 239 AFNTY-NGGMSYGASGVYEF 257
Db 591 AINTSGNKGSYNIGVNYEF 610

```

```

RESULT 2
US-09-336-447A-7
; Sequence 7, Application US/09336447A
; Patent No. 6310190
; GENERAL INFORMATION:
; APPLICANT: HANSEN, ERIC J.
; APPLICANT: AEBI, CHRISTOPH
; APPLICANT: COPE, LESLIE D.
; APPLICANT: MACIVER, ISOBEL
; APPLICANT: FISKE, MICHAEL J.

```


US-09-543-681A-6966
; Sequence 6966, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 6966
; LENGTH: 568
; TYPE: PRT
; ORGANISM: Proteus mirabilis
US-09-543-681A-6966

Query Match 9.4%; Score 125.5; DB 4; Length 568;
Best Local Similarity 26.1%; Pred. No. 8.7e-05;
Matches 42; Conservative 21; Mismatches 49; Indels 49; Gaps 5;

Qy 137 LDARISKKNQIDT-----ISKYLL-----ELGTLYLDDSYRM 168
Db 404 VESDVKNKEDIQTNRNINQVKQTADKNKIYVDKYLANNFDVNEQSKSLVGLSLYAG 463

Qy 169 MEQNTNIN-----KLSKELOTGLANOSALSMVLQV-NGVKTSTVSA 209
Db 464 NKSNTDINAINKDLSHFQNETNRRFYKVKRANQGIASVAAMSNL--PFNDAAFTSTAM 521

Qy 210 AVGYRDKTALAIGVGRITDRFTAKAGVAFNTYNGGMSYG 250
Db 522 GIGNYNATAFAMQYRINENVKNVKAStANDANNWVSAG 562

RESULT 7
US-09-077-098A-6
; Sequence 6, Application US/09077098A
; Patent No. 6544519
; GENERAL INFORMATION:
; APPLICANT: TOKUNAGA, Eiji
; SAKAGUCHI, Masashi
; MATSUO, Kazuo
; HAMADA, Fukusaburo
; TOKIYOSHI, Sachio
; TITLE OF INVENTION: NOVEL POLYPEPTIDE FROM HAEMOPHILUS
; PARAGALLINARUM AND PROCESS FOR PREPARING THE SAME
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 624 Ninth Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: 19-May-1998
; APPLICATION NUMBER: US/09/077,098A
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP97/03222
; FILING DATE: 12-SEP-1997
; APPLICATION NUMBER: JP 27,148/1996
; FILING DATE: 19-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: KORNBAU, Anne M.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: TOKUNAGA=1

US-09-077-098A-6
; Sequence 7, Application US/09077098A
; Patent No. 6544519
; GENERAL INFORMATION:
; APPLICANT: TOKUNAGA, Eiji
; SAKAGUCHI, Masashi
; MATSUO, Kazuo
; HAMADA, Fukusaburo
; TOKIYOSHI, Sachio
; TITLE OF INVENTION: NOVEL POLYPEPTIDE FROM HAEMOPHILUS
; PARAGALLINARUM AND PROCESS FOR PREPARING THE SAME
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 624 Ninth Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: 19-May-1998
; APPLICATION NUMBER: US/09/077,098A
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP97/03222
; FILING DATE: 12-SEP-1997
; APPLICATION NUMBER: JP 27,148/1996
; FILING DATE: 19-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: KORNBAU, Anne M.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: TOKUNAGA=1

TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2042 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-09-077-098A-6

Query Match 8.6%; Score 115; DB 4; Length 2042;
Best Local Similarity 29.8%; Pred. No. 0.009;
Matches 31; Conservative 20; Mismatches 47; Indels 6; Gaps 3;

Qy 157 ELGTLYLDDSYRMMEQNTNINIKLSKELQTLGLANOSALSMVLQVNGVKTSTVSAAVGYRD 216
Db 1942 QLNTVIDNVQNNFNQVNRIGDLTRESRAGIAGMATASLQNVALPGKTTISVGTATPKG 2001

Qy 217 KTLATGVGRITDRFTAKAGV---APTNYNGGMSYGASVGYEF 257
Db 2002 ENAVAIGM-SRLSD--NGKVGIRLSGMSSTNGDKGAAMSVGF 2042

RESULT 8
US-09-077-098A-7
; Sequence 7, Application US/09077098A
; Patent No. 6544519
; GENERAL INFORMATION:
; APPLICANT: TOKUNAGA, Eiji
; SAKAGUCHI, Masashi
; MATSUO, Kazuo
; HAMADA, Fukusaburo
; TOKIYOSHI, Sachio
; TITLE OF INVENTION: NOVEL POLYPEPTIDE FROM HAEMOPHILUS
; PARAGALLINARUM AND PROCESS FOR PREPARING THE SAME
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 624 Ninth Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: 19-May-1998
; APPLICATION NUMBER: US/09/077,098A
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP97/03222
; FILING DATE: 12-SEP-1997
; APPLICATION NUMBER: JP 27,148/1996
; FILING DATE: 19-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: KORNBAU, Anne M.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: TOKUNAGA=1

TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2039 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide

```
;
; SEQUENCE DESCRIPTION: SEQ ID NO: 7:
US-09-077-098A-7

Query Match      8.5%; Score 114; DB 4; Length 2039;
Best Local Similarity 29.8%; Pred. No. 0.012;
Matches 31; Conservative 20; Mismatches 47; Indels 6; Gaps 3;

Qy 157 ELGTYLDDSYRMMEQNTHINKLSKELQTLGLANQSALSMVOPNGVGVKTSVSAAVGGYRD 216
Db 1939 QLVNTVDVNNQVNNQVNRIGDLTRESRAGIAGAVATASLQNVALPGKTTISVGTATFKG 1998

Qy 217 KTLAIGVGSRTIDRTAKAGV---AFNTYNGGMSYGASVGYEF 257
Db 1999 ENAVAIGM-SRLSD--NGKVGIRLSGMSSTSGDKGAAMSVGFTF 2039

RESULT 9
US-09-336-447A-13
; Sequence 13, Application US/09336447A
; Patent No. 6310190
; GENERAL INFORMATION:
; APPLICANT: HANSEN, ERIC J.
; APPLICANT: AEBI, CHRISTOPH
; APPLICANT: COPE, LESLIE D.
; APPLICANT: MACIVER, ISOBEL
; APPLICANT: FISKE, MICHAEL J.
; APPLICANT: FREDENBURG, ROSS A.
; TITLE OF INVENTION: USP41 AND USP42 ANTIGENS OF MORAXELLA CATARRHALIS
; FILE REFERENCE: AMCY:024
; CURRENT APPLICATION NUMBER: US/09/336,447A
; CURRENT FILING DATE: 1999-06-21
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 13
; LENGTH: 873
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-09-336-447A-13

Query Match      8.2%; Score 110; DB 4; Length 873;
Best Local Similarity 27.6%; Pred. No. 0.0083;
Matches 42; Conservative 30; Mismatches 52; Indels 28; Gaps 8;

Qy 110 LGINRQOLKLNLSYFNDLRHDFKLKVLDAISKKNQIDTISKYLLLELGTLYLDDSYRMM 169
Db 745 IATNKOELIQLQ-----ILQHD-RLNRINETNNRQDKID-----QLG-----YALK 781

Qy 170 EQNTHINKLSK-ELQTL--GLANQSALSMVOPNGVGVKTSVSAAVGGYRDKTALAIG-VG 225
Db 782 EQQHFNNRISAVERTAGGIANAIATLPSPSRAGEHHVLFSGSYHNGQAASVLSGAAG 841

Qy 226 SRTIDRTAKAGVAFNTYNGGMSYGASVGYEF 257
Db 842 LSDTGKSTYKIGLSWSD-AGGLSGGVGGSYRW 872

RESULT 10
US-09-336-447A-5
; Sequence 5, Application US/09336447A
; Patent No. 6310190
; GENERAL INFORMATION:
; APPLICANT: HANSEN, ERIC J.
; APPLICANT: AEBI, CHRISTOPH
; APPLICANT: COPE, LESLIE D.
; APPLICANT: MACIVER, ISOBEL
; APPLICANT: FISKE, MICHAEL J.
; APPLICANT: FREDENBURG, ROSS A.
; TITLE OF INVENTION: USP41 AND USP42 ANTIGENS OF MORAXELLA CATARRHALIS
; FILE REFERENCE: AMCY:024
; CURRENT APPLICATION NUMBER: US/09/336,447A
; CURRENT FILING DATE: 1999-06-21
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn Ver. 2.1
```

```
; SEQ ID NO 5
; LENGTH: 892
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-09-336-447A-5

Query Match      8.0%; Score 108; DB 4; Length 892;
Best Local Similarity 27.6%; Pred. No. 0.014;
Matches 42; Conservative 29; Mismatches 53; Indels 28; Gaps 8;

Qy 110 LGINRQOLKLNLSYFNDLRHDFKLKVLDAISKKNQIDTISKYLLLELGTLYLDDSYRMM 169
Db 764 IATNKOELIQLQ-----ND-----RLNRINETNNRQDKID-----QLG-----YALK 800

Qy 170 EQNTHINKLSK-ELQTL--GLANQSALSMVOPNGVGVKTSVSAAVGGYRDKTALAIG-VG 225
Db 801 EQQHFNNRISAVERTAGGIANAIATLPSPSRAGEHHVLFSGSYHNGQAASVLSGAAG 860

Qy 226 SRTIDRTAKAGVAFNTYNGGMSYGASVGYEF 257
Db 861 LSDTGKSTYKIGLSWSD-AGGLSGGVGGSYRW 891

RESULT 11
US-09-336-447A-1
; Sequence 1, Application US/09336447A
; Patent No. 6310190
; GENERAL INFORMATION:
; APPLICANT: HANSEN, ERIC J.
; APPLICANT: AEBI, CHRISTOPH
; APPLICANT: COPE, LESLIE D.
; APPLICANT: MACIVER, ISOBEL
; APPLICANT: FISKE, MICHAEL J.
; APPLICANT: FREDENBURG, ROSS A.
; TITLE OF INVENTION: USP41 AND USP42 ANTIGENS OF MORAXELLA CATARRHALIS
; FILE REFERENCE: AMCY:024
; CURRENT APPLICATION NUMBER: US/09/336,447A
; CURRENT FILING DATE: 1999-06-21
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 831
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-09-336-447A-1

Query Match      8.0%; Score 107; DB 4; Length 831;
Best Local Similarity 27.6%; Pred. No. 0.016;
Matches 42; Conservative 29; Mismatches 53; Indels 28; Gaps 8;

Qy 110 LGINRQOLKLNLSYFNDLRHDFKLKVLDAISKKNQIDTISKYLLLELGTLYLDDSYRMM 169
Db 703 IATNKOELIQLQ-----ND-----RLNRINETNNRQDKID-----QLG-----YALK 739

Qy 170 EQNTHINKLSK-ELQTL--GLANQSALSMVOPNGVGVKTSVSAAVGGYRDKTALAIG-VG 225
Db 740 EQQHFNNRISAVERTAGGIANAIATLPSPSRAGEHHVLFSGSYHNGQAASVLSGAAG 799

Qy 226 SRTIDRTAKAGVAFNTYNGGMSYGASVGYEF 257
Db 800 LSDTGKSTYKIGLSWSD-AGGLSGGVGGSYRW 830

RESULT 12
US-09-540-236-3401
; Sequence 3401, Application US/09540236
; Patent No. 6673910
; GENERAL INFORMATION:
; APPLICANT: GARY L. BRETON ET AL.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATARRHALIS
; FILE REFERENCE: 2709.2005-001
; CURRENT APPLICATION NUMBER: US/09/540,236
```

```

; CURRENT FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 3840
; SEQ ID NO 3401
; LENGTH: 878
; TYPE: prt
; ORGANISM: M.catarrhalis
US-09-540-236-3401

```

Query Match	8.0%;	Score 107;	DB 4;	Length 878;
Best Local Similarity	27.6%;	Pred. No. 0.018;		
Matches	42;	Conservative	29;	Mismatches 53; Indels 28; Gaps 8;

QY	110	LGINRQQLKLNLYSYFNDLRHDFKLKVLDA	RI	SKNKNIDITISKYLELELGYLD	SYRM	169
DB	750	IATNKOELLIQ-----ND-----	KLNRINETNNRQDKID-----	OLG-----	YALK	786
QY	170	EQMTNHNINKLSK-ELQTF--GLANOSAL	SMLVQPNPGVGKTSVSAAVGGVR	KDTALAI	VG	225
DB	787	EQCHFNRRISAVERTQAGGINAAIATAT	PSPSRAGEHHVLFGSGYHNGQAA	VSILGA	AG	846
QY	226	SRITDRPTAKAGVAFNTYNGM	SYGASVGYEF			257
DB	847	LSPTGKSTYKIGLWSPD-AGGLSGVG	SGSYRW			877

```

RESULT 13
US-09-336-447A-9
; Sequence 9, Application US/09336447A
; Patent No. 6310190
; GENERAL INFORMATION:
; APPLICANT: HANSEN, ERIC J.
; APPLICANT: AEBI, CHRISTOPH
; APPLICANT: COPE, LESLIE D.
; APPLICANT: MACIVER, ISOBEL
; APPLICANT: FISKE, MICHAEL J.
; APPLICANT: FREDENBURG, ROSS A.
; TITLE OF INVENTION: USPA1 AND USPA2
; FILE REFERENCE: AMCY:024
; CURRENT APPLICATION NUMBER: US/09/336-447A-9
; CURRENT FILING DATE: 1999-06-21
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 9
; LENGTH: 941
; TYPE: PR1
; ORGANISM: Moraxella catarrhalis
US-09-336-447A-9

```

	Query Match	8.0%;	Score 107;	DB 4;	Length 941;
	Best Local Similarity	25.9%;	Pred. No. 0.02;		
	Matches	41;	Conservative	36;	Mismatches 63;
					Indels 18; Gaps 8;
QY	113	NRQOLKLNLYSYFNDLRHDFKLVLDARISGNQNI----	DTISKYLLLELGTYLDD----	164	
DB	788	NKTAIEONIN--RTVANGFEIERKKGATNKQELILQNDRLNQ--	INETNNHODQKIDQ	843	
QY	165	-SYRMMEQNTHNINKLSK-ELQT--GLANGOSMLVQPNQGVGKTSVSAAVGGVYDKTAL	220		
DB	844	LGVALKEQGHFNRRISAVERTAGGTANAIAITLPSPAGEHHVLFSGSYINGQAAV	903		
QY	221	AIG-VGSRITDRFTAKGAVFNFTYNGMWSYGASVGYEF	257		
DB	904	SICGAAGLSDTGKSTYKIGLSWSD--AGGLSGGVGGSYEW	940		

RESULT 14
US-09-268-347-24
; Sequence 24, Application US/09268347
; Patent No. 6335182
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; TITLE OF INVENTION: RECOMBINANT HAEN
; FILE REFERENCE: 1038-860

```

: CURRENT APPLICATION NUMBER: US/09/268,347
:
: CURRENT FILING DATE: 1999-03-16
:
: NUMBER OF SEQ ID NOS: 54
:
: SOFTWARE: PatentIn ver. 2.0
:
: SEQ ID NO 24
:
: LENGTH: 1002
:
: TYPE: PRT
:
: ORGANISM: Haemophilus influenzae
US-09-268-347-24

```

	Query Match	7.7%	Score 104;	DB 4;	Length 1002;
	Best Local Similarity	36.1%;	Pred.No. 0.047;		
	Matches	30;	Conservative	15;	Mismatches 36;
					Indels 2;
					Gaps 2;
Qy	176	INKLSKELQTLGNQSLMSLWPNVGKTSVSAAVGGYRDKTALATGVGSRTDRTAK	235		
		::::	::::	::::	::::
		::::	::::	::::	::::
Db	921	VNKVGKADAGTASALAASQLPQASMSGKMSVSTAGSSYQGSGLAIGV-SRISDNGKVI	979		
		::::	::::	::::	::::
		::::	::::	::::	::::
Qy	236	AGVAFNTYNGGMS-YGASGVGYEP	257		
		::::	::::	::::	::::
Db	980	IRLSGTTNSOGKTVGAAGVGYOW	1002		
		::::	::::	::::	::::
		::::	::::	::::	::::

```

RESULT 15
US-09-268-347-30
; Sequence 30, Application US/09268347
; Patent No. 6335182
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; TITLE OF INVENTION: RECOMBINANT HAEMOPHILUS INFLUENZAE ADHESIN PROTEINS
; FILE REFERENCE: 1038-860
; CURRENT APPLICATION NUMBER: US/09/268,347
; CURRENT FILING DATE: 1999-03-16
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 30
; LENGTH: 1004
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-09-268-347-30

```

	Query Match	7.7%	Score 104;	DB 4;	Length 1004;
	Best Local Similarity	36.1%	Pred. No. 0.047;		
	Matches	30;	Conservative	15;	Mismatches 36;
					Indels 2;
					Gaps 2;
Qy	176	INKLSKELOQLGLANQSLMSLVDPNGVGKTSVSAAGVGYRDKTALATGVGSRRITDRFTAK	235		
		:::::	:::::	:::::	:::::
Db	919	VNKVGKRADAGTASALAAASQLPQASMPGKSMVSTAGSSYQQSGLAIGV-SRISDNGKVI	977		
		:::::	:::::	:::::	:::::
Qy	236	AGVAFNTYNGGMS-YGASVGYEYF	257		
		:::::	:::::	:::::	:::::
Db	978	IRLSGTTNSQGGKTGVAAGVGYOW	1000		
		:::::	:::::	:::::	:::::

Search completed: May 13, 2004, 08:18:02
Job time : 23 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 13, 2004, 08:16:37 ; Search time 48 Seconds
(without alignments)
1489.858 Million cell updates/sec

Title: US-10-030-529A-2
Perfect score: 1342
Sequence: 1 MKIKCLVAVVGLACSTITM.....VAFNTYNGMSYGASVGYEF 257

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1145568 seqs, 278261457 residues

Total number of hits satisfying chosen parameters: 1145568

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	193	14.4	610	10	US-09-952-267-11
2	193	14.4	624	10	US-09-952-267-7
3	193	14.4	889	10	US-09-952-267-15
4	190.5	14.2	573	10	US-09-952-267-3
5	159.5	11.9	2712	12	US-10-282-122A-67070
6	115	8.6	2042	14	US-10-192-584-6
7	114	8.5	2039	14	US-10-192-584-7
8	110	8.2	873	10	US-09-952-267-13
9	108	8.0	852	12	US-10-282-122A-62892
10	108	8.0	892	10	US-09-952-267-5
11	107	8.0	831	10	US-09-952-267-1
12	107	8.0	941	10	US-09-952-267-9
13	100.5	7.5	1098	9	US-09-797-862-32
14	100.5	7.5	2353	9	US-09-797-862-33
15	97	7.2	496	9	US-09-738-626-5483

16	96.5	7.2	795	12	US-10-282-122A-63527	Sequence 63527, A
17	90.5	6.7	723	15	US-10-369-493-10942	Sequence 10942, A
18	90.5	6.7	819	12	US-10-282-122A-54896	Sequence 54896, A
19	90.5	6.7	3073	12	US-10-282-122A-49147	Sequence 49147, A
20	90	6.7	478	9	US-09-815-242-11674	Sequence 11674, A
21	90	6.7	512	12	US-10-282-122A-59775	Sequence 59775, A
22	90	6.7	1055	14	US-10-032-585-7652	Sequence 7652, Ap
23	89.5	6.7	1074	12	US-10-282-122A-50616	Sequence 50616, A
24	88	6.6	335	12	US-10-282-122A-60846	Sequence 60846, A
25	88	6.6	797	15	US-10-320-800-6	Sequence 6, Appl1
26	88	6.6	1246	12	US-10-282-122A-49773	Sequence 49773, A
27	87	6.5	827	12	US-10-282-122A-60889	Sequence 60889, A
28	86.5	6.4	400	15	US-10-369-493-10982	Sequence 10982, A
29	86.5	6.4	1107	12	US-10-282-122A-75398	Sequence 75398, A
30	86.5	6.4	1225	14	US-10-233-553-20	Sequence 20, Appl1
31	86	6.4	473	12	US-10-282-122A-76706	Sequence 76706, A
32	86	6.4	652	12	US-10-282-122A-49839	Sequence 49839, A
33	85.5	6.4	1461	12	US-10-282-122A-75339	Sequence 75339, A
34	85	6.3	278	12	US-10-424-599-264683	Sequence 264683, A
35	85	6.3	633	12	US-10-282-122A-52615	Sequence 52615, A
36	84.5	6.3	356	9	US-09-771-382-37	Sequence 37, Appl1
37	84.5	6.3	382	9	US-09-771-382-38	Sequence 38, Appl1
38	84.5	6.3	407	9	US-09-771-382-25	Sequence 25, Appl1
39	84.5	6.3	433	9	US-09-771-382-26	Sequence 26, Appl1
40	84.5	6.3	461	9	US-09-771-382-35	Sequence 35, Appl1
41	84.5	6.3	502	9	US-09-771-382-27	Sequence 27, Appl1
42	84.5	6.3	512	9	US-09-771-382-23	Sequence 23, Appl1
43	84.5	6.3	540	9	US-09-771-382-33	Sequence 33, Appl1
44	84.5	6.3	591	9	US-09-797-862-11	Sequence 11, Appl1
45	84.5	6.3	591	9	US-09-797-862-21	Sequence 21, Appl1

ALIGNMENTS

RESULT 1

US-09-952-267-11
; Sequence 11, Application US/09952267
; Publication No. US20030032772A1
; GENERAL INFORMATION:
; APPLICANT: HANSEN, ERIC J.
; APPLICANT: AEBI, CHRISTOPH
; APPLICANT: COPE, LESLIE D.
; APPLICANT: MACIVER, ISOBEL
; APPLICANT: FISKE, MICHAEL J.
; APPLICANT: FREDENBURG, ROSS A.
; TITLE OF INVENTION: USP2A ANTIGENS OF MORAXELLA CATARRHALIS
; FILE REFERENCE: AMCY.024
; CURRENT APPLICATION NUMBER: US/09/952,267
; CURRENT FILING DATE: 2001-09-12
; PRIOR APPLICATION NUMBER: 09/336,447
; PRIOR FILING DATE: 1999-06-21
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 11
; LENGTH: 610
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-09-952-267-11

Query Match	14.4%	Score 193;	DB 10;	Length 610;
Best Local Similarity	37.9%	Pred. No. 1.5e-10;		
Matches 53;	Conservative 18;	Mismatches 61;	Indels 8;	Gaps 3;
Qy	126	NDLRHPFKLVLDARISKQKQNTDTTSKYLLEGTGLDD-SYRMEQNTHT-----NINK 178		
Db	471	NKASDTKFAATADATKGNALTKNAKSIDLGTGVDGFRVTALDTKNAFDGRITA 530		
Qy	179	LSKELQTGLANQASLMLVQPNVGKTSVSAAGVGRDKTALAIQVGSRTIDRETAQGV 238		
Db	531	LDSKVENGAQAALSGLFQPSVGFNATAALGGYGSKAVALGAGYRVPNPLAFKAGA 590		
Qy	239	AFNTY-NGMSYGASVGYEF 257		

Query Match 14.4%; Score 193; DB 10; Length 889;
Best Local Similarity 37.9%; Pred. No. 2.6e-10;


```

;
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/192,584
; FILING DATE: 11-Jul-2002
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/077,098
; FILING DATE: 19-May-1998
; APPLICATION NUMBER: PCT/JP97/03222
; FILING DATE: 12-SEP-1997
; APPLICATION NUMBER: JP 27,148/1996
; FILING DATE: 19-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: KORNBAU, Anne M.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: TOKUNAGA=1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2042 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-10-192-584-6

Query Match      8.6%; Score 115; DB 14; Length 2042;
Best Local Similarity 29.8%; Pred. No. 0.083;
Matches 31; Conservative 20; Mismatches 47; Indels 6; Gaps 3;

QY 157 ELGTYLDDSYRMEQNTNINKLSKELOTLGANSALSMLVQPNVGKTSVSAAVGGYRD 216
DB 1942 QLTNTVIDVNQNFVNQVQRIQDLTRESRAGIAGAMATASLQNVALPKKTTISVGTATFKG 2001

QY 217 KTAALTGVSGRITDRPTAKAGV---AFNTYNGMSYGASVGYEF 257
DB 2002 ENAVAIGM-SRLSD-NGKVGIRLSCWSTSNCGDKGAAMSVGFSF 2042

RESULT 7
US-10-192-584-7
; Sequence 7, Application US/10192584
; Publication No. US20030027987A1
; GENERAL INFORMATION:
; APPLICANT: TOKUNAGA, Eiji
;             SAKAGUCHI, Masaashi
;             MATSUO, Kazuo
;             HAMADA, Fukusaburo
;             TOKIYOSHI, Sachio
; TITLE OF INVENTION: NOVEL POLYPEPTIDE FROM HAEMOPHILUS
;             PARAGALLINARUM AND PROCESS FOR PREPARING THE SAME
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 624 Ninth Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/192,584
; FILING DATE: 11-Jul-2002
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/077,098

```



```

Qy 170 EQWTHNINKLSK-ELQT--GLANQSALSMVLQPNVGKTSVSAAVGGYRDKTALAIG-VG 222
Db 740 EQQGHFNRRISAVERTQAGGIANAIATATLPSPSRAGEHHVLFSGSYHNGQAASVLSGAAG 799
Qy 226 SRITDRTAKAGVAFNTYNGMSYGASVGVEF 257
Db 800 LSDTGKSTYKIGLSWSD-AGGLSGGVGSYRW 830

RESULT 12
US-09-952-267-9
; Sequence 9, Application US/09952267
; Publication No. US2003032772A1
; GENERAL INFORMATION:
; APPLICANT: HANSEN, ERIC J.
; APPLICANT: AEBI, CHRISTOPH
; APPLICANT: COPE, LESLIE D.
; APPLICANT: MACIVER, ISOBEL
; APPLICANT: FISKE, MICHAEL J.
; APPLICANT: FREDENBURG, ROSS A.
; TITLE OF INVENTION: USP1A AND USP2A ANTIGENS OF MORAXELLA CATARRHALIS
; FILE REFERENCE: AMCY.024
; CURRENT APPLICATION NUMBER: US/09/952,267
; CURRENT FILING DATE: 2001-09-12
; PRIOR APPLICATION NUMBER: 09/336,447
; PRIOR FILING DATE: 1999-06-21
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 9
; LENGTH: 941
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-09-952-267-9

Query Match 8.0%; Score 107; DB 10; Length 941;
Best Local Similarity 25.9%; Pred No. 0.18; 63; Indels 18; Gaps 8
Matches 41; Conservative 36; Mismatches 36;

Qy 113 NRQOLKNLYSYFNDLRHDFKLVLDARISKNKNQNI----DTISKYLLLELGTLYLDD---- 164
Db 788 NKTAIEQVIN--RTVANGFEIEKNKAGIATNKQELILQNDRLNQ-INETNNHQDQKIDQ 843
Qy 165 -SYRMEQNTNHNKLSK-ELQT--GLANQSALSMVLQPNVGKTSVSAAVGGYRDKTAL 220
Db 844 LGYALKEQGGHFNNRISAVERTQAGGIANAIATATLPSPSRAGEHHVLFSGSYHNGQAASV 903
Qy 221 AIG-VGSRITDRTAKAGVAFNTYNGMSYGASVGVEF 257
Db 904 SLGAAGLSDTGKSTYKIGLSWSD-AGGLSGGVGSYRW 940

RESULT 13
US-09-797-862-32
; Sequence 32, Application US/09797862
; Patent No. US20020102276A1
; GENERAL INFORMATION:
; APPLICANT: PEAK, IAN RICHARD ANSELM
; APPLICANT: JENNINGS, MICHAEL PAUL
; APPLICANT: MOXON, E. RICHARD
; TITLE OF INVENTION: NOVEL SURFACE ANTIGEN
; FILE REFERENCE: 065064/0134
; CURRENT APPLICATION NUMBER: US/09/797,862
; CURRENT FILING DATE: 2001-05-03
; PRIOR APPLICATION NUMBER: PCT/AU98/01031
; PRIOR FILING DATE: 1998-12-14
; PRIOR APPLICATION NUMBER: GB 9726398.2
; PRIOR FILING DATE: 1997-12-12
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 32
; LENGTH: 1098
; TYPE: PRT
; ORGANISM: Haemophilus influenzae

```


A;Reference number: AB0001; MUID:21470413; PMID:11586360
A;Accession: AH0110
A;Status: preliminary
A:Molecule type: DNA
A;Residues: 1-658 <KUR>
A;Cross-references: GB:AL590842; PIDN:CAC89747.1; PID:g15978974; GSPDB:GN00175
C;Genetics:
A;Gene: YPO0902

Query Match 8.9%; Score 119.5; DB 2; Length 658;
Best Local Similarity 27.4%; Pred.No. 0.1;
Matches 31; Conservative 25; Mismatches 50; Indels 7; Gaps 4;

QY 147 NIDTISKYLELGTLYDDSYRMEONTNINKLSKELOTGLANQSALSMLVOPNGVGKTS 206
 | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 551 NFDOQLKSISNOTNAYTNQRYSLEKQ---DLRKQNSVLGASIASAMWASLTQPYTSGSM 607
 | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
QY 207 VSAAVGGRRDKRTALAIGVGRITD--RFTAKAGVAENTYNGMSYGASGVGEF 257
 | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 608 TTIGAASYRGOSALSILGVS--ISDSGRWSKLQASSNT-QGDFIGVGVGYQIW 658
 | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :

RESULT 7
AE0169
probable exported protein YPO1387 [imported] - Yersinia pestis (strain CO92)
C;Species: Yersinia pestis
C;Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 02-Nov-2001
C;Accession: AE0169
R;Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrall,
Nature 413, 523-527, 2001
A;Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A;Reference number: AB0001; MUID:21470413; PMID:11586360
A;Accession: AE0169
A;Status: preliminary
A:Molecule type: DNA
A;Residues: 1-364 <KUR>
A;Cross-references: GB:AL590842; PIDN:CAC90216.1; PID:g15979436; GSPDB:GN00175
C;Genetics:
A;Gene: YPO1387

Query Match 8.4%; Score 112.5; DB 2; Length 364;
Best Local Similarity 26.8%; Pred.No. 0.16;
Matches 40; Conservative 25; Mismatches 69; Indels 15; Gaps 6;

QY 111 GINRQQKLNLKYFNFLRDHDFLK--VLDAIRSKNKONIDTISKYLLELGTLYLDSDSYM 168
 | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 226 GKSETLN-LNYDIV-----DSRVSSIVDVNSVTDTKVNTAFETSLSDAKSYVDKKY-- 277
 | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
QY 169 MEQTHNINKLSKELOTGLANQSALSMLVOPNGVGKTSVSAAVCGYGRDKRTALAIGVGSRI 228
 | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 278 -NQSLDRKNFKNNFTNAGISGAWSGIPIPKFGYEK-SFGMAICAYRGQALAVGGDWNI 335
 | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
QY 229 TDRETKAGVAENTYNGMSYGASGVGEF 257
 | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 336 NHKTTITRNVSADTEGG---VGVGAAGFAF 361
 | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :

RESULT 8
D90697
adhesin/invasin-like protein [imported] - Escherichia coli (strain O157:H7, substrain RN
C;Species: Escherichia coli
C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C;Accession: D90697
R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gene
A;Reference number: A99629; MUID:21156231; PMID:11258796
A;Accession: D90697
A;Status: preliminary
A:Molecule type: DNA

A;Residues: 1-338 <HAY>
A;Cross-references: GB:BAO00007; PID:NABJ33971.1; PID:g13360006; GSPDB:GN00154
A;Experimental source: strain O157:H7, substrain RMD 050952
C;Genetics:
A;Gene: ECs0548

Query Match 8.2%; Score 109.5; DB 2; Length 338;
Best Local Similarity 20.1%; Pred. No. 0.26;
Matches 56; Conservative 32; Mismatches 92; Indels 99; Gaps 9

Qy 73 TYLELOHYMPYTPVLVLYAPGVSPSP-----ILLYPMSDDP 108
 ||| ||| :||
Db 65 TYLEHHYIPSETDPTTQTPTVQDPAGQKTAATGDVTARTYQSMINARQSASVTDQA 124
 | | | | :

Qy 109 QLGINRROOLKU-----NLYSY-----FNDLR 129
 ||| ||| ::
Db 125 QTOITEQQAQIVATOKTLAATGDTQNTAHYOEMINARLAAQNEANORTATEQGCKNNALT 184
 | | | | :

Qy 130 HDFKLKVLDAISKNKO-----NIDITSKYLLELTGLTLDSDSRMWEONTHNI 176
 ||| ||| :||
Db 185 TDVAVOQNERTQYDKQMQLSQAESAQAHEQLDSLSQDVTTQHOOI/TWTKRVADNSQOI 244
 | | | | :

Qy 177 NKLS-----KELOTGLANQSALSMLVQPNGVGKTS----VSAAVGGYRDKT 218
 ||| ||| :||
Db 245 NTLNNHFSSLKNEVDNRKANAGTA--SAIAIASOPQ--VKTGDMVMVSAGAGTFNCES 300
 | | | | :

Qy 219 ALAIGVGSIRTRDTAKGAFVNITYNGMSGYSVGYEYF 257
 ||| ||| :||
Db 301 AVSVGTGFNACTHTVLKAGISADT-QSDFGAGVGVGYSF 338
 | | | | :

RESULT 9
G85547
adhesin/invasin-like protein [similarity] - Escherichia coli (strain O157:H7,
C:Species: Escherichia coli
C>Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 02-Nov-2001
C:Accession: G85547
R.;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.
iller, L.; Grobetz, E.J.; Davis, N.W.; Lim, A.; Dimailanta, E.; Potamousis, K.
Nature 409, 529-533, 2001
A>Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: G85547
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-338 <STO>
A:Cross-references: GB:AEO05174; NID:g12513376; PID:N.AAG54843.1; GSPDB:GN00145
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A;Gene: Z0639

Query Match 8.2%; Score 109.5; DB 2; Length 338;
Best Local Similarity 20.1%; Pred. No. 0.26;
Matches 56; Conservative 32; Mismatches 92; Indels 99; Gaps 9

Qy 73 TYLELOHYMPYTPVLVLYAPGVSPSP-----ILLYPMSDDP 108
 ||| ||| :||
Db 65 TYLEHHYIPSETDPTTQTPTVQDPAGQKTAATGDVTARTYQSMINARQSASVTDQA 124
 | | | | :

Qy 109 QLGINRROOLKU-----NLYSY-----FNDLR 129
 ||| ||| ::
Db 125 QTOITEQQAQIVATOKTLAATGDTQNTAHYOEMINARLAAQNEANORTATEQGCKNNALT 184
 | | | | :

Qy 130 HDFKLKVLDAISKNKO-----NIDITSKYLLELTGLTLDSDSRMWEONTHNI 176
 ||| ||| :||
Db 185 TDVAVOQNERTQYDKQMQLSQAESAQAHEQLDSLSQDVTTQHOOI/TWTKRVADNSQOI 244
 | | | | :

Qy 177 NKLS-----KELOTGLANQSALSMLVQPNGVGKTS----VSAAVGGYRDKT 218
 ||| ||| :||
Db 245 NTLNNHFSSLKNEVDNRKANAGTA--SAIAIASOPQ--VKTGDMVMVSAGAGTFNCES 300
 | | | | :

Qy 219 ALAIGVGSIRTRDTAKGAFVNITYNGMSGYSVGYEYF 257
 ||| ||| :||

Db 301 AVSGVTSFNAGTHTVLKAGISADT-QSDFGAGVGVGSF 338

RESULT 10
AF0169
probable exported protein YP01388 [imported] - Yersinia pestis (strain C092)
C:Species: Yersinia pestis
C>Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 02-Nov-2001
C:Accession: AF0169
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.; deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.; Hill, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrrell, Nature 413, 523-527, 2001
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A:Reference number: AB0001; MUID:21470413; PMID:11586360
A:Accession: AF0169
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-622 <KUR>
A:Cross-references: GB:AL590842; PIDN:CAC90217.1; PID:gl5979437; GSPDB:GNO0175
C:GeneID:81
A:Gene: YP01388

Query Match 7.6%; Score 102.5; DB 2; Length 622;
Best Local Similarity 25.0%; Pred. No. 2.2;
Matches 37; Conservative 31; Mismatches 63; Indels 17; Gaps 6;

Qy 113 NRQQLKLNLYSFNDLRHDFKLVLDARISKKNQIDT-ISKYLL-----ELGTYYLDDSYR 167
Db 485 NLIDVTLLNANNYTDLS-----VNTIYYTGKQYTDTSRINEYQRTFKNEFLTYSNGKFG 537
Qy 168 MMEQNTNHNKLSKELOTGLANQSALSMVLQPNVGKTSVAAGVGYRDKTALAIAGVGR 227
Db 538 GFDRK---DINQKQQLNAGTAATMAAAVTPQKSG-SKVSIGVGLAGYSQGGAGVGAHW 593
Qy 228 ITRFTAKAGVAFNTYNG-QMSYGASVG 254
Db 594 VNQRITMNTTMTYDTQRGVSLLTGLSIG 621

RESULT 11
S76235
hypothetical protein - Synechocystis sp. (strain PCC 6803)
C:Species: Synechocystis sp.
A:Variety: PCC 6803
C>Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 02-Feb-2001
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; O, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda DNA Res. 3, 109-136, 1996
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis sp.
A:Reference number: S74322; MUID:97061201; PMID:8905231
A:Accession: S76235
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-831 <KAN>
A:Cross-references: EMBL:D90914; GB:AB001339; NID:gl653477; PIDN:BAA18494.1; PID:dl01922
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C:Keywords: ATP; nucleotide binding; p-loop
F:317-324/Region: nucleotide-binding motif A (p-loop)
F:409-414/Region: nucleotide-binding motif B
F:413-416/Region: DEXH motif

Query Match 7.4%; Score 99.5; DB 2; Length 831;
Best Local Similarity 20.1%; Pred. No. 5.8;
Matches 52; Conservative 32; Mismatches 70; Indels 105; Gaps 9;

Qy 48 SNEGGFDI-----KVPGIKMKPREWISKQ--ATYLE-----LQHYMPYTPVLVTVYAGVSP 96
Db 167 SKEGDYSYRLQRKQVGLTKALRDLLKRWVESLYLEPPKAVQPOLPKQVTLNFDP---- 222
Qy 97 SPILLYPMSDPDQLGINROQLKLNLYSFNDRHDFKLVLDAR---ISKKNQIDTISK 153

Db 223 -----SFKLTQLQARTLVLKQGG----- 242

Qy 154 YLLELGTYYLDDSYRMEQNTHNINKLSKELOTGLANQSALSMVLQPN----- 200

Db 243 -----EMWLADLLKAVPCSASTIOSLAKKGLVALAREKRLRFOQPSINASOAPELTPAQ 297

Qy 201 -----GVGKTSVSAAGVGYR---DKTALA-----IGVGSRIIDR 231

Db 298 KQACQTVLPLOQYHQVLLHGVYGGSKTEVYLIQICDRLRGKQSVLVLVPELGLTPQLTDR 357

Qy 232 FTAKAGVAFNTYNGMSYG 250

Db 358 FRARFGNKVAVVYHGLSSG 376

RESULT 12
A56921
kinesin family protein KIF1a - mouse
C:Species: Mus musculus (house mouse)
C>Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 02-Feb-2001
C:Accession: A56921
R:Okada, Y.; Yamazaki, H.; Sekine-Aizawa, Y.; Hirokawa, N. Cell 81, 769-780, 1995
A:Title: The neuron-specific kinesin superfamily protein KIF1a is a unique monomeric motor
A:Reference number: A56921; MUID:95292344; PMID:7539720
A:Accession: A56921
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1695 <RES>
A:Cross-references: GB:D29951; NID:g976234; PIDN:BAA06221.1; PID:g976235
C:Superfamily: kinesin-related protein unc-104; kinesin motor domain homology; pleckstrin
C:Keywords: nucleotide binding; p-loop
F:6-360/Domain: kinesin motor domain homology <KNOT>
F:97-104/Region: nucleotide-binding motif A (p-loop)

Query Match 7.3%; Score 97.5; DB 2; Length 1695;
Best Local Similarity 24.1%; Pred. No. 23;
Matches 47; Conservative 34; Mismatches 71; Indels 43; Gaps 10;

Qy 4 KCLVAVGLACSTTTTMAQQPKFAGVSSLYSYEYDYGKRWTSNEGFDIKVPGIKMK 63
Db 25 KCIIQMSG---STTIVNPKQPKETPKS--FSFDYSY-----MSHTSPEDI----- 65

Qy 64 KPEWISKQATYLE-----LQH-YMPYTPVLVTVYAGVSPSPILLYPMSDPDQLGINROQL 117

Db 66 --NYASQKQVYRDIGEEMLQHAFGYGVNVCIFAYGQTGAGKSYTMGKQKQGGI-IFQL 122

Qy 118 KLNLYSYFNDLRHD-----FKLVLDARISKKNQIDTISKYLLLELGTYYLDD 164

Db 123 CEDLESRINDTNDNMSYSVEVSMEIYCERVDLLNPKNGNL-RVREHPL-LGPLYVED 180

Qy 165 SYRMEQNTHNINKL 179

Db 181 LSKLAVTSYNDIQDL 195

RESULT 13
A95179
hypothetical protein Sp1538 [imported] - Streptococcus pneumoniae (strain TIGR4)
C:Species: Streptococcus pneumoniae
C>Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 03-Aug-2001
C:Accession: A95179
R:Petzelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heidon, J.D.; Unayan, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapfel, nson, T.; Hickey, E.K.; Holt, I.E. Science 293, 498-506, 2001
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison, A:Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.
A:Reference number: A95000; MUID:21357209; PMID:11463916
A:Accession: A95179
A:Status: preliminary
A:Molecule type: DNA

A;Residues: 1-466 <KUR>
A;Cross-references: GB:AE05672; PIDN:AAK75626.1; PID:g14973027; GSPDB:GN00164; TIGR:SP4
A;Experimental source: strain TIGR4
C;Genetics:
A;Gene: SPI538

Query Match	7.2% ¹	Score 97	DB 2	Length 466
Best local Similarity	24.7%	Pred. No. 4.2		
Matches	59	Conservative 30	Mismatches 88	Indels 62
				Gaps 14
Qy	45	WTWSNEGGFDIKVP----	GIKMKPKWISKQATVLELQHYMPYTPVLVTVYAPGV-----	94
Db	156	WTFEDKGD-DLHLPDSLSDLKRWVW-----	HOHSSDIVPISGSKATGVKVKVEHL	205
Qy	95	--SPSPKILLYPMGSDPDOLGINRQOLKUNLVXSYFN-----	DLRHDFKVLVDARISKNNQN	147
Db	206	GLXPEKVMVF--GD-----GLN-----	DLELFDYAGISVAMGISHD-----	KIKEK 244
Qy	148	IDRTISKVLLLELGTLYLD--DSVYRMEQNTH-----	NINKLSKELQTLQANQSALSMVLVOPNGV	202
Db	245	ADVIYTKLEEDGIFAALVEFGVMEKEHLFPQVDIETVEGFLATIKTNHGDRLKLPFEHA		304
Qy	203	GKT----SVSAAVGGYRDKTALAIGVGSRIITDRFTAKAGVAFNTYNGMS--YGASVGVEF		257
Db	305	PKTVANFVLSKDGYYDGV-----IFHRIIKDFMIQCGDPTGTGCMGESITYGGSFDEF		358

RESULT 14
D64226
endopeptidase La (EC 3.4.21.53) - Mycoplasma genitalium
N;Alternate names: ATP-dependent proteinase lon; ATP-dependent serine proteinase La
N;Contains: adenosinetriphosphatase (EC 3.6.1.3)
C;Species: Mycoplasma genitalium
C;Date: 09-Apr-1999 #sequence_revision 09-Apr-1999 #text_change 02-Jun-2003
C;Accession: D64226
R;Fraser, C.M.; Gocayne, J.D.; White, O.; Adams, M.D.; Clayton, R.A.; Fleischmann, R.D.;
M.; Fuhrmann, J.; Nguyen, D.; Utterback, T.R.; Saudek, D.M.; Phillips, C.A.; Merrick, J.
C.A.; Venter, J.C.
Science 270, 397-403, 1995
A;Title: The minimal gene complement of Mycoplasma genitalium.
A;Reference number: A64200; MUID:96026346; PMID:7569993
A;Accession: D64226
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-795 <TIGR>
A;Cross-references: GB:I39702; GB:I43967; NID:g1045928; PID:g1045929; TIGR:MG239
A;Experimental source: strain G-37
C;Comment: This enzyme catalyzes the hydrolysis of large proteins in the presence of ATP
C;Genetics:
A;Gene: lon
A;Genetic code: SGC3
C;Superfamily: ATP-dependent Lon protease
C;Keywords: ATP; DNA binding; heat shock; hydrolase; nucleotide binding; P-loop; serine
F;379-386/Region: nucleotide-binding motif A (P-loop)
F;442-447/Region: nucleotide-binding motif B
F;702/Active site: Ser #status predicted

[illegible]

Qy 189 --NQSALSMVLVQPNGVGKTSVSAAVGGYRDKTALAIGVG 225
| : | | | | | : : : | : | : |
Db 368 QKNKGPIMLVGPVGKSSLAKEALDKGFVKISLG 406

RESULT 15

H98045
hypothetical protein sprl393 [imported] - Streptococcus pneumoniae (strain R6)
C:Species: Streptococcus pneumoniae
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 22-Oct-2001
C:Accession: H98045
R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszczak, L.; Burgett, S.; DeHoff
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAh
V, P.; Sun, P.M.; Winkler, M.E
J. Bacteriol. 183, 5709-5711, 2001
A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jasku
A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A:Reference number: A97872; MUID:21429245; PMID:11544234
A:Accession: H98045
A>Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-466 <KUR>
A:Cross-references: GB:AE007317; PIDN:AAL00197.1; PID:gi15459043; GSPDB:GN00174
C:Genetics:
A:Gene: sprl393

Query Match	7.1%	Score 95;	DB 2;	Length 466;
Best Local Similarity	24.74;	Pred. No. 6.1;		
Matches 59; Conservative		30; Mismatches	88; Indels	62; Gaps 14;
Qy	45	WTWSNEGGEFDIKVP----	GIKMKPEWISKQATYLELQHYMPYTPVLVTPYAGV-----	94
Db	156	WTPEDKGD-DLHLFPDSLSDLKVRW----	-----HQHSSDIVIPSGSKATGEVKVVEHL	205
Qy	95	--SPSPILLVPMGSDPDOLGINROOLKUNLVXSYFN-----	DLRHDFKULVDLARIKKNQKN	147
Db	206	GLKPEKVMWF--GD-----GLN-----	DLELFDYAGISVAMGISHD-----	KIKEK 244
Qy	148	IDTISKYLLLELGY-LDDSVRMMEQNTH-----	NINKLSKELQTLGANSALSMVLQPNGV	202
Db	245	ADVIITKLEEDGIFDALEVFGEVKEKELHFPOVDIETVEGFP	PLATIKTNHGDLRIKLFFEHA	304
Qy	203	GKT-----SVSAAVGGYRDKTALAIGVGSRIITDR	TAKAGAVFNTYNGMS--YGASVGYEF	257
Db	305	PKTVANFVLSLKQGYVDGV-----IFHRIITKDFMI	OGGDDPTGTGMGGESYIGHSFDEF	358

Search completed: May 13, 2004, 08:16:00
Job time : 22 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 13, 2004, 08:14:11 ; Search time 17 Seconds
(without alignments)
787.178 Million cell updates/sec

Title: US-10-030-529a-2
Perfect score: 1342
Sequence: 1 MKIKLVAVVGLACSTITM.....VAFNTYNGMSYGASGVGEF 257

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	186.5	13.9	455	1	YADA YEREN
2	182	13.6	434	1	YADA YEREN
3	99.5	7.4	831	1	PRIA_SYNY3
4	99	7.4	1690	1	KPIA_HUMAN
5	97.5	7.3	1695	1	KPIA_MOUSE
6	96.5	7.2	795	1	LON_MYCGE
7	92.5	6.9	588	1	CMC2_CABEL
8	92.5	6.9	727	1	MPPI_ARATH
9	90.5	6.7	819	1	LON_CHLPN
10	90.5	6.7	2716	1	OSA_DROME
11	90	6.7	1328	1	POLX_TOBAC
12	89.5	6.7	956	1	UVRA_LISIN
13	88.5	6.6	498	1	FLID_BACSU
14	88	6.6	4543	1	LRPI_CHICK
15	87	6.5	806	1	LON_FORBU
16	86.5	6.4	524	1	CLPF_CHLRE
17	86.5	6.4	1226	1	DSRA_HUMAN
18	86	6.4	473	1	Y085_UREPA
19	86	6.4	766	1	TKL1_MOUSE
20	86	6.4	920	1	YML1_YEAST
21	85.5	6.4	218	1	GTH5_ARATH
22	85.5	6.4	522	1	LEU1_SHEON
23	85	6.3	677	1	YD64_MYCPN
24	85	6.3	810	1	HPUB_NEIMC
25	84	6.3	491	1	G6PD_BUCAI
26	84	6.3	662	1	YME1_SCHMA
27	84	6.3	1011	1	PERO_MOUSE
28	83.5	6.2	763	1	PEPX_STRMC
29	83.5	6.2	817	1	PERQ_HUMAN
30	83.5	6.2	2214	1	POLG_CXA24
31	83	6.2	290	1	CYOA_BUCAP
32	83	6.2	518	1	GSH1_BUCAP
33	82.5	6.1	436	1	TOLB_RHIME

Query Match 13.9%; Score 186.5; DB 1; Length 455;
Best Local Similarity 34.1%; Pred. No. 6.3e-08;

ALIGNMENTS

RESULT 1

YADA YEREN
ID YADA_YEREN STANDARD; PRT; 455 AA.
AC P31489;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Invasin precursor [Outer membrane adhesin].
GN YADA OR YOPA OR INVA OR YOP1.
OS Yersinia enterocolitica.
OG Plasmid pIV.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Yersinia.
OX NCBI_TaxID=630;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=6471/76 / Serotype O:3;
RX MEDLINE=95020586; PubMed=7934875;
RA Tamm A., Tarkkanen A., Korhonen T.K., Kuusela P., Toivanen P.,
RA Skurnik M.;
RT "Hydrophobic domains affect the collagen-binding specificity and
RT surface polymerization as well as the virulence potential of the Yada
RT protein of Yersinia enterocolitica.";
RL Mol. Microbiol. 10:995-1011(1993).
CC -!- FUNCTION: INVASIN IS A PROTEIN THAT ALLOWS ENTERIC BACTERIA TO
CC PENETRATE CULTURED MAMMALIAN CELLS. THE ENTRY OF INVASIN IN THE
CC CELL IS MEDIATED BY BINDING SEVERAL BETA-1 CHAIN INTEGRINS. THIS
CC PROTEIN IS AN ADHESIN FORMING A FIBRILLAR MATRIX ON THE CELL
CC SURFACE.
CC -!- SUBCELLULAR LOCATION: Outer membrane.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X13882; CAA32086.1; -.
DR PIR; S04912; S04912.
DR InterPro; IPR008126; Adhesion.
DR InterPro; IPR008640; Hep_Hag.
DR InterPro; IPR008635; HIM.
DR InterPro; IPR005594; Yada.
DR Pfam; PF05658; Hep_Hag; 4.
DR Pfam; PF05662; HIM; 1.
DR Pfam; PF03895; Yada; 1.
DR PRINTS; PR01756; OMADHESIN.
DR Plasmid; Virulence; Signal; Outer membrane.
FT SIGNAL 1 25
FT CHAIN 26 455
SQ SEQUENCE 455 AA; 47136 MW; AC12EF68C657DAC0 CRC64;

34 82.5 6.1 438 1 MURD_YERPE Q8zif1 yersinia pe
35 82.5 6.1 824 1 HELI_HSV6U P52356 human herpe
36 82.5 6.1 824 1 HELI_HSV6Z P52450 human herpe
37 82.5 6.1 956 1 UVRA_LISMO Q8y4f6 listeria mo
38 82.5 6.1 1748 1 YNR2_YEAST P53886 saccharomyc
39 82.5 6.1 1876 1 GLS1_YEAST P38631 saccharomyc
40 81.5 6.1 313 1 ENGC_RHOBA Q7uuz6 rhodospirell
41 81.5 6.1 359 1 OMPA_SERMA P04845 serratia ma
42 81.5 6.1 384 1 YZ43_METJA Q60298 methanococc
43 81.5 6.1 826 1 CRAA_BACUH Q96597 bacillus th
44 81 6.0 276 1 OCCT_AGRTU P35121 agrobacteri
45 81 6.0 366 1 APNI_YEAST P22936 saccharomyc

Matches	47; Conservative	25; Mismatches	57; Indels	9; Gaps	4;
QY	120 NLYSYFNDLRHFKLVLDARISKNNQNIDITISKYLLLEGLTYLDDSYRMMEQNTHINKL	179			
Db	327 NVYA---DKSSHTLKTANSYTDVTVSN--STKKAIRESNQYTDHKEFQLD--NRLLDKL	378			
QY	180 SKELQTLGNQASALSMILVQPNGVKTSVSAAVGGYRDKTLAIAIGVSGRIITDRFTTKAGVA	239			
Db	379 DTRVDKGLASSAALNSLFQPVGVGVNFTAGVGGYRSSQALAISSGYRVNENVALKAGVA	438			
QY	240 FNTYNGMSYGSVGYEFP	257			
Db	439 Y-AGSSDVMYNASFNIEW	455			

RESULT 2

YADA_ID	YADA_YERPS	STANDARD;	PRT;	434 AA.
AC	PI0858;			
DD	01-JUL-1989	(Rel. 11, Created)		
DT	01-JUL-1989	(Rel. 11, Last sequence update)		
DT	01-JUL-1993	(Rel. 26, Last annotation update)		
DE	Invasin precursor	(Outer membrane adhesin).		
DE	YADA OR YOPA OR INVA OR YOP1.			
GN	Yersinia pseudotuberculosis.			
OG	Plaemid pibi.			
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;			
OC	Enterobacteriaceae; Yersinia.			
OX	NCBI_TaxID=633;			
[1]	RN			
RC	SEQUENCE FROM N.A.			
RC	STRAIN-YPIII;			
RX	MEDLINE=88302441; PubMed=3043229;			
RT	Rosqvist R., Skurnik M., Wolf-Watz H.;			
RA	"Increased virulence of Yersinia pseudotuberculosis by two			
RT	independent mutations."			
RL	Nature 334:522-525(1988).			

[2] RN SEQUENCE FROM N.A.
RP STRAIN=YPIII;
RC MEDLINE=69343638; PubMed=2761389;
RX Skurnik M., Wolf-watz H.;
RT "Analysis of the yopA gene encoding the YopI virulence determinants of *Yersinia* spp.;
RL Mol. Microbiol. 3:517-529(1989).

NOTES: MICROBIAL: 93-217-3253A13937.

-1- SURCELLULAR LOCATION: Outer membrane
SURFACE.

--- SUBCELLULAR LOCATION: Outer membrane.

CC This SWISS-PROT entry is copyright. It is produced through a collabora-
CC tion between the Swiss Institute of Bioinformatics and the EMBL outst-
CC

CC the European Bioinformatics Institute. There are no restrictions (CC
CC use by non-profit institutions as long as its content is in the
CC

CC modified and this statement is not removed. Usage by and for com
CC entities requires a license agreement (See <http://www.fish-sib.ch/ans>
CC

CC BY-NC-ND 4.0 International license. For more information, see <http://creativecommons.org/licenses/by-nc-nd/4.0/> or send an email to license@isb-sib.ch.

CC

DR EMBL; X13883; CAA32088.1; -.

DR PIR; S04534; S04534.
DR InterPro; IPR008126; Adhesion.

DR	InterPro; IPR008640; Hep_Hag.
DB	InterPro; IPR008635; HIM

DR InterPro; IPR008635; HM.
DR InterPro; IPR005594; Yada.
DR Pfam; PF050. Yada; Yada.

DR Pfam; PF05658; Hep_Hag; 4.
DR Pfam; PF05662; HIM; 1.

DR pfam; PF03895; YadaA; 1.
DR PRINTS; PR01756; OMADHESIN.

KW	Plasmid; Virulence; Signal; Outer membrane.
FT	SIGNAL. 1 25

FT	SIGNAL	1	23	
FT	CHAIN	26	434	INVASIN.

```

SQ      SEQUENCE      434 AA;  45054 MW;  EE2C55F8B12B183D4  CRC64;

Query Match      13.6%;  Score 182;  DB 1;  Length 434;
Best Local Similarity 39.0%;  Pred. NO. 1.4e-07;
Matches 41;  Conservative 19;  Mismatches 41;  Indels 4;  Gaps 2;

QY      153 KYLLELGYLDSDYRMEQNTNINKLSKELTGLANQASMLVQPNVGKTSVSAAVG 212
      - - - - - : : : : : : : : : : : : : : : : : : : : : :
Db      334 KAISESNQYTDHKFSQLD--NRLDKLDRVDKGLASSAALNSLFQYGVGKVFAGVG 390

QY      213 GYRDKTALAIGVRSITDRFTAKAGVAFNTYNGMSYSGASVGEF 257

Db      391 GYRSSQALAIIGSYRVNESVALKAGVAY-AGSSNVMYNFASFRIEW 434

```

RESULT 3

```

PRIA_SYN3
ID PRIA_SYN3 STANDARD; PRT; 831 AA.
AC P74397;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DT Primosomal protein N' (Replication factor Y).
DE PRIA OR SLL0270.
GN Synnechocystis sp. (strain PCC 6803).
OS Bacteria; Cyanobacteria; Chroococcales; Synnechocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97061201; PubMed=8905231;
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hiroseawa M., Sugiura M., Sasamoto S., Kimura T.,
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,
RA Shimpou S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
RA Tabata S.;
RT "sequence analysis of the genome of the unicellular cyanobacterium
RT Synnechocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions.";
RL DNA Res. 3:109-136(1996).
CC -!- FUNCTION: Recognizes a specific hairpin sequence on phiX ssDNA.
CC This structure is then recognized and bound by proteins priB and
CC priC. Formation of the primosome proceeds with the subsequent
CC actions of dnaB, dnaC, dnaI and primase. priA then functions as a
CC helicase within the primosome (By similarity).
CC -!- SIMILARITY: Belongs to the helicase family. priA subfamily.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/
CC or send an email to license@isb-sib.ch).

```

00 OF SEND AN EMAIL TO LICENSE@ISD.SID.CN/.

DR EMBL; D90914; BAA18494.1; -.
DR PIR; S76235; S76235.

DR InterPro: IPR001410; DEAD.
DR InterPro: IPR001650; Helicase C.

```
DR  interPro; IPR001030; release_C.  
DR  interPro; IPR005259; PriA.  
DR  Pfam; PF00370; DEAD_1
```

```
DR pfam; PF00370; DEAD; 1.
DR pfam; PF00271; helicase_C; 1.
```

DR SMART; SM00487; DEXDC; 1.
DR SMART; SM00490; HELICC; 1.

DR TIGRFAMS; TIGR00595; priA; 1.
KW DNA replication: DNA-binding: ATP-binding: Helicase: Primosome:

KW Zinc-finger; Complete proteome.
 MW DNA replication; DNA-binding; AIF-binding; necroptosis; ribosome;
 ND BIND 317 324 AND (POTENTIAL)

FT	NP BIND	317	324	ATP (POTENTIAL) .
FT <td>SITE</td> <td>413</td> <td>416</td> <td>DEEH BOX.</td>	SITE	413	416	DEEH BOX.

FT	ZN FING	537	C4-TYPE (POTENTIAL).
FT	ZN FING	549	C4-TYPE (POTENTIAL).
FT	ZN FING	568	C4-TYPE (POTENTIAL).

SEQ	SEQUENCE	831 AA;	93306 MW;	EA52CDF95654D94A CRC64;
1	1	1	1	1
2	2	2	2	2
3	3	3	3	3
4	4	4	4	4
5	5	5	5	5
6	6	6	6	6
7	7	7	7	7
8	8	8	8	8
9	9	9	9	9
10	10	10	10	10
11	11	11	11	11
12	12	12	12	12
13	13	13	13	13
14	14	14	14	14
15	15	15	15	15
16	16	16	16	16
17	17	17	17	17
18	18	18	18	18
19	19	19	19	19
20	20	20	20	20
21	21	21	21	21
22	22	22	22	22
23	23	23	23	23
24	24	24	24	24
25	25	25	25	25
26	26	26	26	26
27	27	27	27	27
28	28	28	28	28
29	29	29	29	29
30	30	30	30	30
31	31	31	31	31
32	32	32	32	32
33	33	33	33	33
34	34	34	34	34
35	35	35	35	35
36	36	36	36	36
37	37	37	37	37
38	38	38	38	38
39	39	39	39	39
40	40	40	40	40
41	41	41	41	41
42	42	42	42	42
43	43	43	43	43
44	44	44	44	44
45	45	45	45	45
46	46	46	46	46
47	47	47	47	47
48	48	48	48	48
49	49	49	49	49
50	50	50	50	50
51	51	51	51	51
52	52	52	52	52
53	53	53	53	53
54	54	54	54	54
55	55	55	55	55
56	56	56	56	56
57	57	57	57	57
58	58	58	58	58
59	59	59	59	59
60	60	60	60	60
61	61	61	61	61
62	62	62	62	62
63	63	63	63	63
64	64	64	64	64
65	65	65	65	65
66	66	66	66	66
67	67	67	67	67
68	68	68	68	68
69	69	69	69	69
70	70	70	70	70
71	71	71	71	71
72	72	72	72	72
73	73	73	73	73
74	74	74	74	74
75	75	75	75	75
76	76	76	76	76
77	77	77	77	77
78	78	78	78	78
79	79	79	79	79
80	80	80	80	80
81	81	81	81	81
82	82	82	82	82
83	83	83	83	83
84	84	84	84	84
85	85	85	85	85
86	86	86	86	86
87	87			

Query Match 7.4%; Score 99.5; DB 1; Length 831;

```
Best Local Similarity 20.1%; Pred. No. 2;
Matches 52; Conservative 32; Mismatches 70; Indels 105; Gaps 9;

Qy 48 SNEGFDI-----KVPGLKMPKWKISKQ--ATYLE-----LQHMYPYTPVLVTVAPGVSP 96
Db 167 SKEGDYSYRLQKVPGLTKALRDLKQWVSYLEPKAVQPOLPKWVTLNFDP-----222
Qy 97 SPILLYPMSDDQLGINRQQLKLNLYSYFNDLRHDFKLKVLDAK-----ISKKNQKIDITISK 153
Db 223 -----SFKLTELOARTLVLKKNQGG-----242
Qy 154 YLLELGTLYDPSYRMMEQNTNINKLSKELQTLGLANQALSMLVQPN-----200
Db 243 -----EWLADLLKAVPCSASTIQSLAKKGLVAIAEREKRLFFQQPSINASQAPELTPAQ 297
Qy 201 -----GVGKTSVSAAGVYR---DKTALA-----LGVGSRTYDR 231
Db 298 KQACQTVLPQGYHQVLLHGVTSKGTEVYLQICDRLGKQGSVLVLVPEIGLTPQLTDR 357
Qy 232 FTAKAGVAFNTYNGMSYG 250
Db 358 FRFRGNKVAVTHSLSG 376

RESULT 4
KF1A HUMAN
ID KF1A HUMAN STANDARD; PRT; 1690 AA.
AC Q12756;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Kinesin-like protein KIF1A (Axonal transporter of synaptic vesicles).
GN ATSV OR KIF1A.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=96299637; PubMed=8661001;
RA Furlong R.A., Zhou C.Y., Ferguson-Smith M.A., Affara N.A.;
RT "Characterization of a kinesin-related gene ATSV, within the tuberosus
RT sclerosis locus (TSC1) candidate region on chromosome 9q34.";
RL Genomics 33:421-429(1996).
CC -!- FUNCTION: Motor for anterograde axonal transport of synaptic
CC vesicle precursors (By similarity).
CC -!- SUBUNIT: Monomer (By similarity).
CC -!- SIMILARITY: Belongs to the kinesin-like protein family. Unc-104
CC subfamily.
CC -!- SIMILARITY: Contains 1 FHA domain.
CC -!- SIMILARITY: Contains 1 PH domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X90840; CAA62346.1; .
CC HSSP; P17119; 3KAR.
CC Genew; HGNC:888; ATSV.
CC MIW; 601255; .
CC GO; GO:0003774; P: motor activity; TAS.
CC GO; GO:0008089; P: anterograde axon cargo transport; TAS.
CC InterPro; IPR000253; FHA.
CC InterPro; IPR001752; kinesin_motor.
CC InterPro; IPR001849; PH.
CC InterPro; IPR008984; SMAD_FHA.
CC Pfam; PF00498; FHA; 1.
CC Pfam; PF00225; Kinesin; 1.
```

```
Pfam; PF00169; PH; 1.
DR PRINTS; PR00380; KINESINHEAVY.
DR SMART; SM00240; FHA; 1.
DR SMART; SM00129; KISC; 1.
DR SMART; SM00233; PH; 1.
DR PROSITE; PS00411; KINESIN MOTOR DOMAIN1; 1.
DR PROSITE; PS0067; KINESIN MOTOR DOMAIN2; 1.
DR PROSITE; PS0006; FHA DOMAIN; 1.
DR PROSITE; PS0003; PH DOMAIN; 1.
KW Motor protein; Microtubule; ATP-binding; Coiled coil.
FT DOMAIN 1 361 KINESIN-MOTOR.
FT DOMAIN 366 383 COILED COIL (POTENTIAL).
FT DOMAIN 429 462 COILED COIL (POTENTIAL).
FT DOMAIN 516 572 FHA.
FT DOMAIN 622 681 COILED COIL (POTENTIAL).
FT DOMAIN 801 822 COILED COIL (POTENTIAL).
FT DOMAIN 1575 1673 PH.
FT NP BIND 97 104 ATP (POTENTIAL).
SQ SEQUENCE 1690 AA; 191083 MW; D8DDEC784624FB4D CRC64;

Query Match 7.4%; Score 99; DB 1; Length 1690;
Best Local Similarity 24.1%; Pred. No. 5.5;
Matches 53; Conservative 34; Mismatches 83; Indels 50; Gaps 11;

Qy 4 KCLVAVVGLACSTTTTMAQQPPKPAFAGVSSLYSYEYDYGKGTWNSGEGFDIKVPGIKMK 63
Db 25 KCIQMSG---STTTIYNPKQKPTPKS--FSPDYSY-----WSHTSPEDI-----65
Qy 64 PKWISKQATYLE-----LQH-YMPYTPVLVTVAPGVSPSPILLYPMSDDQLGINRQOL 117
Db 66 --NYASQKQVYRDIGEMQLQHAPEGVNVCIFAYGTGAGKSYTMWGKQKQOGI-IPQL 122
Qy 118 KLNLYSFPNLRHD-----FKLVLDARISKKNQKIDITISKYLELLEGTLYDD 164
Db 123 CEDLFSSINDTTNDNMSYSVEVSYMEYICERVRDLNPKNGNL-RVREHPL-LGPYVED 180
Qy 165 -----SYRMMEQNTNINKLSKELQTLGLANQALSMLV 197
Db 181 LSKLAVTSYNDIQDLMDSGNKAARTVAATNNETSRSRAV 220

RESULT 5
KF1A MOUSE
ID KF1A MOUSE STANDARD; PRT; 1695 AA.
AC P33173; Q61770;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Kinesin-like protein KIF1A (Axonal transporter of synaptic vesicles).
GN ATSV OR KIF1A OR KIF1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=95292344; PubMed=7539720;
RA Okada Y., Yamazaki H., Sekine-Aizawa Y., Hirokawa N.;
RT "The neuron-specific kinesin superfamily protein KIF1A is a unique
RT monomeric motor for anterograde axonal transport of synaptic vesicle
RT precursors.";
RL Cell 81:769-780(1995).
RN [2]
RP PRELIMINARY SEQUENCE OF 100-247 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=93077686; PubMed=1447303;
RA Aizawa H., Sekine Y., Takemura R., Zhang Z., Nangaku M.,
RA Hirokawa N.;
RT "Kinesin family in murine central nervous system.";
RL J. Cell Biol. 119:1287-1296(1992).
CC -!- FUNCTION: Motor for anterograde axonal transport of synaptic
CC vesicle precursors.
```

```
CC -1- SUBUNIT: Monomer.
CC -1- TISSUE SPECIFICITY: EXPRESSED ALMOST EXCLUSIVELY IN ADULT BRAIN
CC TISSUE (MAINLY IN THE CEREBELLUM AND CEREBRUM) WITHIN A SINGLE
CC TYPE OF NEURONAL CELL. WITHIN THE NEURONAL CELL LEVELS ARE
CC CONCENTRATED AROUND THE AXON, WITH SMALLER AMOUNTS IN THE
CC PERINUCLEAR AND SYNAPTIC REGIONS.
CC -1- SIMILARITY: Belongs to the kinesin-like protein family. Unc-104
CC subfamily.
CC -1- SIMILARITY: Contains 1 FHA domain.
CC -1- SIMILARITY: Contains 1 PH domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D29951; BAA06221.1; -.
CC PIR; A56921; A56921.
CC DR PDB; 1I58; 30-MAY-01.
CC DR PDB; 1I61; 30-MAY-01.
CC MGD; MGI:108391; Kif1a.
CC DR InterPro; IPR000253; FHA.
CC DR InterPro; IPR001752; kinesin_motor.
CC DR InterPro; IPR001849; PH.
CC DR InterPro; IPR008984; SNAD_FHA.
CC DR Pfam; PF00498; FHA; 1.
CC DR Pfam; PF00225; kinesin; 1.
CC DR Pfam; PF00169; PH; 1.
CC DR SMART; SM00380; KINESINHEAVY.
CC DR SMART; SM00240; FHA; 1.
CC DR SMART; SM00129; KISC; 1.
CC DR SMART; SM00233; PH; 1.
CC DR PROSITE; PS00411; KINESIN MOTOR DOMAIN1; 1.
CC DR PROSITE; PS00067; KINESIN MOTOR DOMAIN2; 1.
CC DR PROSITE; PS50006; FHA DOMAIN; 1.
CC DR PROSITE; PS50003; PH DOMAIN; 1.
CC Motor protein; Microtubule; ATP-binding; Coiled coil; 3D-structure.
KW DOMAIN 1 361 KINESIN-MOTOR.
FT DOMAIN 366 383 COILED COIL (POTENTIAL).
FT DOMAIN 429 462 COILED COIL (POTENTIAL).
FT DOMAIN 516 572 FHA.
FT DOMAIN 622 681 COILED COIL (POTENTIAL).
FT DOMAIN 801 822 COILED COIL (POTENTIAL).
FT DOMAIN 1580 1678 PH.
FT NP BIND 97 104 ATP (BY SIMILARITY).
SQ SEQUENCE 1695 AA; 191724 MW; D6EC3B88CBC9CC6 CRC64;

Query Match 7.3%; Score 97.5; DB 1; Length 1695;
Best Local Similarity 24.1%; Pred. No. 7.3;
Matches 47; Conservative 34; Mismatches 71; Indels 43; Gaps 10;

Qy 4 KCLVAVVGLACSTTTTMAQPPKAGVSSLYSEYDYGKGKWTWSNBSGGFDIKVPGIKMK 63
Db 25 KCIITQMSG---STTTVNPQPKETPKS--FSPDYSY-----WSHTSPEDI----- 65

Qy 64 KPEWISKQATYLE-----LQH-YNPTPLVTVYAPGVSPSILLYPMSDDPQLGINRQQL 117
Db 66 --NYASQKVYRDIGEEMLQHAFEGYNNVICFAYQGTGAGKSYTMMGKQEKDQGGI-IPQL 122

Qy 118 KLNLYSYFNDLRHP-----FKLKVLDAISKKNQKIDTISKVLLLEGTGLDD 164
Db 123 CEDLFSRINDTNDNMSYSVEVSYNEIYCRVRLNLPKNKGNL-RVREHPL-LGPYVD 180

Qy 165 SYRMMEQNTHNINKL 179
Db 181 LSKLAVTSYNDIQDL 195
```

RESULT 6
LON_MYCGE

```
ID LON MYCGE STANDARD; PRT; 795 AA.
AC P47481; Q49276;
DT 01-FEB-1996 (Rel. 33; Created)
DT 01-FEB-1996 (Rel. 33; Last sequence update)
DT 28-FEB-2003 (Rel. 41; Last annotation update)
DE ATP-dependent protease La (EC 3.4.21.53).
GN LON OR MG219.
OS Mycoplasma genitalium.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2097;
RN [1]
RC SEQUENCE FROM N.A. / G-37;
RX STRAIN=ATCC 33530 / G-37;
RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
RA Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,
RA Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhrmann J.L.,
RA Nguyen D.T., Utterback T.R., Saudek D.M., Phillips C.A., Merrick J.M.,
RA Tomb J.-F., Dougherty B.A., Bott K.F., Hu P.-C., Lucier T.S.,
RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;
RT "The minimal gene complement of Mycoplasma genitalium.";
RL Science 270:397-403(1995).
RN [2]
RC SEQUENCE OF 484-606 FROM N.A.
RX STRAIN=ATCC 33530 / G-37;
RA Peterson S.N., Hu P.-C., Bott K.F., Hutchison C.A. III;
RT "A survey of the Mycoplasma genitalium genome by using random
sequencing.";
RL J. Bacteriol. 175:7918-7930(1993).
CC -1- FUNCTION: DEGRADES SHORT-LIVED REGULATORY AND ABNORMAL PROTEINS
CC IN PRESENCE OF ATP. DEGRADES THE REGULATORY PROTEINS RCSA AND
CC SULA. HYDROLYZES TWO ATPs FOR EACH PEPTIDE BOND CLEAVED IN THE
CC PROTEIN SUBSTRATE (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: Hydrolysis of large proteins such as globin,
CC casein and denaturated serum albumin, in presence of ATP.
CC -1- SUBUNIT: Homotetramer (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: Belongs to peptidase family S16.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U39703; AAC71460.1; -.
CC EMBL; U02148; AAD12428.1; -.
CC PIR; D64226; D64226.
CC MEROPS; S16.004; -.
CC TIGR; MG239; -.
CC InterPro; IPR003593; AAA_ATPase.
CC InterPro; IPR003959; AAA_ATPase_cent.
CC InterPro; IPR008269; Pept_S16_C.
CC InterPro; IPR004815; Pept_S16_Lon.
CC InterPro; IPR003111; Pept_S16_N.
CC InterPro; IPR008268; Peptid_S16_AS.
CC InterPro; IPR001984; Peptidase_S16.
CC Pfam; PF00004; AAA; 1.
CC Pfam; PF02190; LON; 1.
CC Pfam; PF05362; Lon_C; 1.
CC PRINTS; PR00830; ENDOLAPTASE.
CC SMART; SM00382; AAA; 1.
CC SMART; SM00464; LON; 1.
CC TIGRFAMs; TIGR00763; Lon; 1.
CC PROSITE; PS01046; LON_SER; 1.
KW Hydrolase; Serine protease; ATP-binding; Complete proteome.
FT DOMAIN 304 308 POLY-SER.
FT NP BIND 379 386 ATP (POTENTIAL).
FT ACT_SITE 702 702 BY SIMILARITY.
FT CONFLICT 494 494 T -> S (IN REF. 2).
```

FT	TRANSMEM	307	326	2 (POTENTIAL).
FT	TRANSMEM	352	365	3 (POTENTIAL).
FT	TRANSMEM	403	422	4 (POTENTIAL).
FT	TRANSMEM	446	463	5 (POTENTIAL).
FT	TRANSMEM	504	523	6 (POTENTIAL).
FT	CA_BIND	52	63	EF-HAND 1.
FT	CA_BIND	86	97	EF-HAND 2.
FT	DOMAIN	123	133	ANCESTRAL CALCIUM SITE 3.
FT	CA_BIND	153	164	EF-HAND 4.
FT	REPEAT	246	332	SOLCAR 1.
FT	REPEAT	342	428	SOLCAR 2.
FT	REPEAT	440	529	SOLCAR 3.
SEQ	SEQUENCE	588 AA;	66336 MW;	68DDF60923D8697D CRC64;

Query Match 6.9%; Score 92.5; DB 1; Length 588;
 Best Local Similarity 22.4%; Pred.No.5;
 Matches 48; Conservative 41; Mismatches 72; Indels 53; Gaps 12

QY	49	NEGGFDIKVGIKMKPKEMISKQATYLEOHMPYTPVLVYAGVSPSPILLVPMSPD	108
DB	89	NDGTIDIR-----DLTLAKHETHP-----ANLAPVIMSKMS-PD	124
QY	109	QLGINQQKLNLVSYFN-DLRHDFKLVDLARISKKNQN-IDTI--SKYLLELGTLYDD	164
DB	125	DEG-----RVDYFSSVYLENEQKLAEMFADMRNHDGLVDVVMKMYCKDIGVPLDD	178
QY	165	SYRMBEQNTNINKSKELQTLGANSALS-----MLVQPNGVGKTSVSAAGVGYRDKTAL	220
DB	179	-----HKAQHVNNMD-----QTGSASVDLKEFFBPMILYPS-----SDLKDIIVDFWRHNLII	226
QY	221	AIGVGRITDRFT---AKAGVAFTNYNGMSYGA	251
DB	227	DIGEDSQIPEDFSQEQMEOGIIWWRHLVAGGAAGA	260

RESULT 8

MFPI_ARATH	STANDARD;	PRT;	727 AA.
AC	Q9LW85;		
DT	16-OCT-2001 (Rel. 40, Created)		
DT	16-OCT-2001 (Rel. 40, Last sequence update)		
DT	16-OCT-2001 (Rel. 40, Last annotation update)		
DE	MAR binding filament-like protein 1.		
GN	MFPI OR AT3G16000 OR MSL1.4.		
OS	Arabidopsis thaliana (Mouse-ear cross).		
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
OC	Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots; rosids;		
OC	eurosid II; Brassicales; Brassicaceae; Arabidopsids.		
NCBI_TaxID=3702;			
[1]	SEQUENCE FROM N.A.		
RP	STRAIN=cv. Columbia;		
RC	MEDLINE=20277480; PubMed=10819329;		
RA	Sato S., Nakamura Y., Kaneko T., Katoh T., Asamizu E., Tabata S.;		
RT	"Structural analysis of Arabidopsis thaliana chromosome 3. I. Sequence		
RT	features of the regions of 4,504,864 bp covered by sixty P1 and TAC		
RT	clones.";		
RL	DNA Res. 7:131-135(2000).		
CC	-!- FUNCTION: Binds DNA. Interacts with chromatin via matrix		
CC	attachment regions (MARs). Likely to participate in nuclear		
CC	architecture by connecting chromatin with the nuclear matrix and		
CC	potentially with the nuclear envelope (By similarity).		
CC	-!- SUBCELLULAR LOCATION: Nuclear matrix (By similarity).		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation		
CC	the-European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See http://www.isb-sib.ch/announcement		
CC	or send an email to license@isb-sib.ch).		
CC	-----		
DR	EMBL; AB012247; BAB02666.1; -.		

RESULT 10
OSA_DROME
ID OSA_DROME STANDARD; PRT; 2716 AA.
AC O81N94; O61603; Q9VEG7;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Trithorax group protein OSA (Eyelid protein).
GN OSA OR ELD OR CG7467.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A., FUNCTION, SUBCELLULAR LOCATION, TISSUE
RP SPECIFICITY, AND DEVELOPMENTAL STAGE.
RX MEDLINE=97415319; PubMed=9271118;
RA Treisman J.E., Luk A., Rubin G.M., Heberlein U.;
RT "eyelid antagonizes wingless signaling during Drosophila development
and has homology to the Bright family of DNA-binding proteins.";
RL Genes Dev. 11:1949-1962 (1997).
RN [2]
RP SEQUENCE FROM N.A.
RP STRAIN=Berkley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.J., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Bailew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista A.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Folsler C., Gabrieli A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.D., Harvey D.A., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D.A., Lai Z.,
RA Laoko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster";
RL Science 287:2185-2195 (2000).
RN [3]
RP FUNCTION, AND DEVELOPMENTAL STAGE.
RX MEDLINE=99112962; PubMed=9895321;
RA Vazquez M., Moore L., Kennison J.A.;
RT "The trithorax group gene osa encodes an ARID-domain protein that
genetically interacts with the brahma chromatin-remodeling factor to
regulate transcription";
RT Development 126:733-742 (1999).

RNA DNA-BINDING, AND IDENTIFICATION IN A BRAHMA COMPLEX WITH BRM AND SNR1.
RX MEDLINE=20069333; PubMed=10601025;
RA Collins R.T., Furukawa T., Tanese N., Treisman J.E.;
RT "Osa associates with the Brahma chromatin remodeling complex and
promotes the activation of some target genes";
RL EMBO J. 18:7029-7040 (1999).
RN [5]
RP FUNCTION
RX MEDLINE=99403006; PubMed=10471712;
RA Saehling-Hampton K., Ciampa P.J., Brook A., Dyson N.;
RT "A genetic screen for modifiers of E2F in Drosophila melanogaster";
RL Genetics 153:275-287 (1999).
RN [6]
RP IDENTIFICATION IN A BRAHMA COMPLEX WITH BRM; OSA; MOR; SNR1; DALAO;
RP BAP55; BAP60 AND BAP47, AND FUNCTION AS COACTIVATOR.
RX MEDLINE=20270023; PubMed=10809665;
RA Kal A.J., Mahmoudi T., Zak N.B., Verrijzer C.P.;
RT "The Drosophila brahma complex is an essential coactivator for the
trithorax group protein zeste";
RL Genes Dev. 14:1058-1071 (2000).
RN [7]
RP FUNCTION AS A COREPRESSOR.
RX MEDLINE=20573925; PubMed=11124806;
RA Collins R.T., Treisman J.E.;
RT "Osa-containing Brahma chromatin remodeling complexes are required
for the repression of wingless target genes";
RL Genes Dev. 14:3140-3152 (2000).
RN [8]
RP FUNCTION AS A COREPRESSOR, AND INTERACTION WITH PNR AND CHI.
RX MEDLINE=22515897; PubMed=12629041;
RA Heitzler P., Vanollet L., Biryukova I., Ramain P.;
RT Enhancer-promoter communication mediated by Chip during
Pannier-driven proneural patterning is regulated by Osa";
RL Genes Dev. 17:591-596 (2003).
RN [9]
RP FUNCTION: Trithorax group (trxG) protein required for embryonic
segmentation, development of the notum and wing margin, and
photoreceptor differentiation. Required for the activation of
genes such as Antp, Ubx and Eve. Binds to DNA without specific
affinity, suggesting that it is recruited to promoters by
promoter-specific proteins. Essential component of the Brahma
complex, a multiprotein complex which is the equivalent of the
yeast SWI/SNF complex and acts by remodelling the chromatin by
catalyzing an ATP-dependent alteration in the structure of
nucleosomal DNA. This complex can both serve as a transcriptional
coactivator or corepressor, depending on the context. Acts as an
essential coactivator for Zeste, which recruits the whole complex
to specific genes. In contrast, it acts as a corepressor for Wg
target genes, possibly via an interaction with Pan and Gro. It
also acts as a negative regulator for proneural achaete-scute,
when it is directly recruited by Pan and Chi. Also represses E2f
activation.
-1- SUBUNIT: Component of the Brahma complex, which is composed of
Brm, Osa, Mor, Snr1/Bap45, Bap11/Dalao, Bap55, Bap60 and Bap47.
Interacts with Pnr and Chi via its EHD domain.
-1- SUBCELLULAR LOCATION: Nuclear.
-1- TISSUE SPECIFICITY: Ubiquitously expressed in early embryo. In
third instar larvae, it is ubiquitously expressed in wing and eye-
antenna imaginal disks, with a stronger expression in a band just
anterior to the morphogenetic furrow.
-1- DEVELOPMENTAL STAGE: Expressed both maternally and zygotically.
-1- DOMAIN: The ARID domains mediates both the binding to DNA.
-1- SIMILARITY: Contains 1 ARID domain.
-1- SIMILARITY: Contains 1 EHD (Eyelid homology) domain.
-1- CAUTION: Ref.2 (AAF55457) sequence differs from that shown due to
erroneous gene model prediction.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

or send an email to license@iab-sib.ch).

```
CC CC -----
DR EMBL; AF053091; AAC06254.1; -.
DR EMBL; AE003718; AAF55457.1; ALT_SEQ.
DR EMBL; AE003718; AAN13750.1; -.
DR PIR; T13049; T13049.
DR FlyBase; FBGN003013; osa.
DR GO; GO:0005634; C:nucleus; IDA.
DR GO; GO:0003677; F:DNA binding; IDA.
DR GO; GO:0046530; P:photoreceptor cell differentiation; IMP.
DR GO; GO:0045449; P:regulation of transcription; IDA.
DR GO; GO:0007379; P:segment specification; IMP.
DR GO; GO:0008587; P:wing margin morphogenesis; IMP.
DR GO; GO:0016055; P:Wnt receptor signaling pathway; IMP.
DR InterPro; IPR001506; ARID.
DR InterPro; IPR008938; ARM.
DR InterPro; IPR006031; XYPPX.
DR Pfam; PF01388; ARID; 1.
DR Pfam; PF02162; XYPPX; 8.
DR SMART; SM00501; BRIGHT; 1.
KW Transcription regulation; DNA-binding; Activator; Repressor;
KW Chromatin regulator; Nuclear protein; Developmental protein.
FT DOMAIN 997 1111
FT DOMAIN 1769 2517
FT DOMAIN 19 1763 PRO-RICH.
FT DOMAIN 174 380 GLN-RICH.
FT DOMAIN 619 873 GLY-RICH.
FT DOMAIN 1222 1453 GLY-RICH.
FT DOMAIN 1271 1751 GLN-RICH.
FT DOMAIN 1730 1745 HIS-RICH.
FT DOMAIN 2589 2624 SER-RICH.
FT DOMAIN 2625 2716 ALA-RICH.
FT CONFLICT 61 61 MISSING (IN REF. 1).
FT CONFLICT 1169 1169 V -> G (IN REF. 1).
FT CONFLICT 1795 1795 M -> T (IN REF. 1).
FT CONFLICT 2637 2637 G -> E (IN REF. 1).
SQ SEQUENCE 2716 AA; 284063 MW; BFAE76CB51C7C675 CRC64;
```

Query Match 6.7%; Score 90.5; DB 1; Length 2716;

Best Local Similarity 23.1%; Pred. No. 50; Mismatches 86; Indels 43; Gaps 8;

Matches 46; Conservative 24; Mismatches 86; Indels 43; Gaps 8;

QY 67 WISKQATYLEQHYMPYTPVLVYAPGVSPSPILLYPMSDPQLGINRQKLNLVSYFN 126

Db 1006 WLDKLRAPWEER---TP--ITACTISKQPLDIY-----RLYIVK 1042

QY 127 DLRHDFLKLVDARISKKNQIDITISKYLELGTLYDDSYRMEQNTNINKLSKELOTG 186

Db 1043 E-RGGF-----VEVTKSKTWKDIAG--LLGIGASSAAAYTLRKHYYTKNLLTFECHFDRG 1093

QY 187 LANQSALSMVQPNVGKTSVSAV-----GGYRDKTALAIGVGSRTIDRFTAKAGVAFNT 242

Db 1094 DIDPLFIQQVEAGSKKTKAKAASVPSPGGHLDAGTTNSTGSSNSQDSFPAPPGSAPNA 1153

QY 243 -----YNGGMSYGASVG 254

Db 1154 AIDGPGPGGSPYPVAGS 1172

RESULT 11

POLX TOBAC

ID POLX TOBAC STANDARD; PRT; 1328 AA.

AC P10978;

DT 01-JUL-1989 (Rel. 11, Created)

DT 01-JUL-1989 (Rel. 11, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Retrovirus-related Pol polyprotein from transposon TNT 1-94 [Contains:

DE Protease [EC 3.4.23.-]; Reverse transcriptase [EC 2.7.7.49];

DE Endonuclease].

OS Nicotiana tabacum (Common tobacco).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;

OC Lamiids; Solanales; Solanaceae; Nicotiana.

```
OX NCBI_TaxID=4097;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89097311; PubMed=2536143;
RA Grandbastien M.-A., Spielmann A., Caboche M.;
RT "Tnt1, a mobile retroviral-like transposable element of tobacco
RL Nature 337:376-380(1989).
CC -! CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate
CC + (DNA) (N).
CC -! SIMILARITY: HIGH, WITH DROSOPHILA COPIA ELEMENT.
CC -! SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY A11.
CC -! SIMILARITY: CONTAINS 1 CCHC-TYPE ZINC FINGER.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/
CC or send an email to license@iab-sib.ch).
CC -----
DR EMBL; X13777; CAA32025.1; -.
DR PIR; S04273; S04273.
DR MEROPS; A11.002; -.
DR InterPro; IPR001995; Peptidase_A2.
DR InterPro; IPR001584; Rve.
DR InterPro; IPR001878; Znf_CCHC.
DR Pfam; PF00665; rve; 1.
DR Pfam; PF00077; rvp; 1.
DR Pfam; PF00098; zf-CCHC; 1.
DR PRINTS; PR00939; C2HCZNFINGER.
DR SMART; SM00343; Znf_C2HC; 1.
DR PROSITE; PSS0158; ZF_CCHC; 1.
KW Hydrolase; Aspartyl protease; RNA-directed DNA polymerase;
KW Endonuclease; Transferase; Polyprotein; Transposable element;
KW Zinc-finger.
FT ZN_FING 230 247 CCHC-TYPE.
FT ACT_SITE 297 297 PROTEASE (BY SIMILARITY).
SQ SEQUENCE 1328 AA; 151076 MW; F2E76C504B19B1B CRC64;
```

Query Match 6.7%; Score 90; DB 1; Length 1328;

Best Local Similarity 24.0%; Pred. No. 22;

Matches 47; Conservative 31; Mismatches 78; Indels 40; Gaps 9;

QY 59 GIKMKPEWTSKQATYLEQHYMPYTPVLVYAPGVSPSPILLYPMSDPQLGINRQ-- 116

Db 962 GLKQAPRQWYMKFDSFMKSQTYLKYSDPCVYFKRFSNNFIILLYYDDMLIVGDKGL 1021

QY 117 -LKLNLVSYFNDLRHDFLK-----VLDARISKKNQIDITISKYLELGTLYDDSYRM 168

Db 1022 IAKLK-----GDLSEKSDMKDLGPAQILGMKIVRER---TSRKLWLSQEKYIE---RV 1069

QY 169 MEQ-NTNINKLSKELOTGLANQSALSMVQPNVG-----GKTSVSAVAGGY----- 214

Db 1070 LERENMKN---AKPVSTFLAGHLKLSKMKCPTTVEERKGNMAKVPYSSAVGSLMAMVCT 1125

QY 215 RDKTALAIGVGSRTID 230

Db 1126 RPDIAHAVGVVSRFLE 1141

RESULT 12

UVRA LISIN

ID UVRA LISIN STANDARD; PRT; 956 AA.

AC Q928A5;

DT 15-MAR-2004 (Rel. 43, Created)

DT 15-MAR-2004 (Rel. 43, Last sequence update)

DT 15-MAR-2004 (Rel. 43, Last annotation update)

DE UVRA ABC system protein A (UvrA protein) (Excinuclease ABC subunit A).

GN Listeria OR Lin2631.

OS Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.


```
FT DOMAIN 2472 2512 EGF-LIKE 10. CLASS A 11. BY SIMILARITY.
FT DOMAIN 2516 2557 LDL-RECEPTOR CLASS A 11. BY SIMILARITY.
FT DOMAIN 2558 2596 LDL-RECEPTOR CLASS A 12. BY SIMILARITY.
FT DOMAIN 2597 2635 LDL-RECEPTOR CLASS A 13. BY SIMILARITY.
FT DOMAIN 2636 2684 LDL-RECEPTOR CLASS A 14. BY SIMILARITY.
FT DOMAIN 2688 2730 LDL-RECEPTOR CLASS A 15. BY SIMILARITY.
FT DOMAIN 2730 2769 LDL-RECEPTOR CLASS A 16. BY SIMILARITY.
FT DOMAIN 2770 2812 LDL-RECEPTOR CLASS A 17. BY SIMILARITY.
FT DOMAIN 2814 2853 LDL-RECEPTOR CLASS A 18. BY SIMILARITY.
FT DOMAIN 2854 2897 LDL-RECEPTOR CLASS A 19. BY SIMILARITY.
FT DOMAIN 2900 2938 LDL-RECEPTOR CLASS A 20. BY SIMILARITY.
FT DOMAIN 2939 2978 EGF-LIKE 11. CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 2979 3019 EGF-LIKE 12.
FT DOMAIN 3020 3058 EGF-LIKE 13.
FT DOMAIN 3059 3097 LDL-RECEPTOR CLASS A 21.
FT DOMAIN 3098 3136 LDL-RECEPTOR CLASS A 22.
FT DOMAIN 3137 3175 LDL-RECEPTOR CLASS A 23.
FT DOMAIN 3176 3214 LDL-RECEPTOR CLASS A 24.
FT DOMAIN 3215 3253 LDL-RECEPTOR CLASS A 25.
FT DOMAIN 3254 3292 LDL-RECEPTOR CLASS A 26.
FT DOMAIN 3293 3331 LDL-RECEPTOR CLASS A 27.
FT DOMAIN 3332 3370 LDL-RECEPTOR CLASS A 28.
FT DOMAIN 3371 3409 LDL-RECEPTOR CLASS A 29.
FT DOMAIN 3410 3448 LDL-RECEPTOR CLASS A 30.
FT DOMAIN 3449 3487 LDL-RECEPTOR CLASS A 31.
FT DOMAIN 3488 3526 EGF-LIKE 14.
FT DOMAIN 3527 3565 EGF-LIKE 15.
FT DOMAIN 3566 3604 EGF-LIKE 16.
FT DOMAIN 3605 3643 EGF-LIKE 17.
FT DOMAIN 3644 3682 EGF-LIKE 18.
FT DOMAIN 3683 3721 EGF-LIKE 19.
FT DOMAIN 3722 3760 EGF-LIKE 20.
FT DOMAIN 3761 3799 EGF-LIKE 21.
FT DOMAIN 3800 3838 EGF-LIKE 22.
FT DOMAIN 3839 3877 RECOGNITION SITE FOR PROTEOLYTICAL
FT DOMAIN 3878 3916 PROCESSING (POTENTIAL).
FT DOMAIN 3917 3955 ENDOCYTOSIS SIGNAL (POTENTIAL).
FT SITE 4001 4506 BY SIMILARITY.
FT DISULFID 29 42 BY SIMILARITY.
FT DISULFID 36 55 BY SIMILARITY.
FT DISULFID 49 66 BY SIMILARITY.
FT DISULFID 74 87 BY SIMILARITY.
FT DISULFID 81 100 BY SIMILARITY.
FT DISULFID 94 110 BY SIMILARITY.
FT DISULFID 117 126 BY SIMILARITY.
FT DISULFID 122 135 BY SIMILARITY.
FT DISULFID 137 150 BY SIMILARITY.
FT DISULFID 156 166 BY SIMILARITY.
FT DISULFID 162 175 BY SIMILARITY.
FT DISULFID 177 190 BY SIMILARITY.
FT DISULFID 190 495 BY SIMILARITY.
FT DISULFID 491 506 BY SIMILARITY.
FT DISULFID 508 521 BY SIMILARITY.
FT DISULFID 805 816 BY SIMILARITY.
FT DISULFID 812 825 BY SIMILARITY.
FT DISULFID 827 840 BY SIMILARITY.
FT DISULFID 852 864 BY SIMILARITY.
FT DISULFID 859 877 BY SIMILARITY.
FT DISULFID 871 888 BY SIMILARITY.
FT DISULFID 893 905 BY SIMILARITY.
FT DISULFID 900 918 BY SIMILARITY.
FT DISULFID 912 929 BY SIMILARITY.
FT DISULFID 934 946 BY SIMILARITY.
FT DISULFID 941 959 BY SIMILARITY.
FT DISULFID 953 969 BY SIMILARITY.
FT DISULFID 974 987 BY SIMILARITY.
FT DISULFID 982 1000 BY SIMILARITY.
FT DISULFID 994 1009 BY SIMILARITY.
FT DISULFID 1013 1025 BY SIMILARITY.
FT DISULFID 1020 1038 BY SIMILARITY.
FT DISULFID 1032 1049 BY SIMILARITY.
FT DISULFID 1060 1073 BY SIMILARITY.
FT DISULFID 1067 1086 BY SIMILARITY.
FT DISULFID 1080 1095 BY SIMILARITY.

Query Match 6.6% Score 88; DB 1; Length 4543;
Best Local Similarity 20.8%; Pred No. 1.5e+02;
Matches 56; Conservative 44; Mismatches 93; Indels 76; Gaps 12;

Qy 29 GVSSLYEYDYGKGTWNSGEGFDI-----KVPQIKMKPEWI 68
Db 1959 GIGRVEGIADWIAAGNIYWDQ-GFDVIEARLNGSRVYVVISQGLDKPRAITVHPEK-- 2015
Qy 69 SKQATYLELQHYMPYTPVLTVPYAGVSPSPILLYPMSPDQOLGINRQOLKLNLYSYFNDL 128
Db 2016 ---GYLFWTEWQYPRIERSRLDGTERRMVLNVNVSISWPNGISVDYEDGKL-----YWCDA 2067
Qy 129 RHDPKLVLDARISKKNQ-----NIDTISKYLELGTLYLDDSYRMMEQNTN---NINKL 179
Db 2068 RTD-KIERIDLETGENREVLSNDMDMFSVFPEDYIYWSN-----RTHANGSIKRG 2119
Qy 180 SKE-----LQTL-----ANQSALSMVLQPNQGVKTSVSAAVGGYRDKTAL 220
Db 2120 SKDNATESVSLRTGIGVQLKDIKVFNRARQKGTNVCAQNG-GCQQLCLFRGGGRRTAC 2178
Qy 221 AIGVGSRIITRFTAKAGVAFNTYNGMSY 249
Db 2179 AHGM-----LSEdGVSCRDYDGLLY 2199

RESULT 15
LON BORBU STANDARD; PRT; 806 AA.
AC Q59185;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE ATP-dependent protease La (EC 3.4.21.53).
GN LON OR BB0253.
OS Borrelia burgdorferi (Lyme disease spirochete).
OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia.
OX NCBI_TaxID=139;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 35210 / B31;
RX MEDLINE=97409968; PubMed=9266683;
RA Cloud J.L., Marconi R.T., Eggers C.H., Garon C.F., Tilly K.,
Samuels D.S.;
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 13, 2004, 08:14:11 ; Search time 45 Seconds
(without alignments)
1801.960 Million cell updates/sec

Title: US-10-030-529A-2

Perfect score: 1342

Sequence: 1 MKIKCLVAVVGLACSTITM.....VAFTYNGMSYGASGVGEF 257

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 25: *
1: sp_archaea: *
2: sp_bacteria: *
3: sp_fungi: *
4: sp_human: *
5: sp_invertebrate: *
6: sp_mammal: *
7: sp_mhc: *
8: sp_organellae: *
9: sp_phage: *
10: sp_plant: *
11: sp_rodent: *
12: sp_virus: *
13: sp_vertebrate: *
14: sp_unclassified: *
15: sp_rvirus: *
16: sp_bacteriap: *
17: sp_archaeap: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1342	100.0	257	16 Q9K2H6	Q9K2H6 haemophilus
2	1299.5	96.8	264	2 Q9K3B8	Q9K3B8 haemophilus
3	1289	96.1	271	2 Q9L916	Q9L916 haemophilus
4	1283	95.6	273	2 Q9L917	Q9L917 haemophilus
5	1262	94.0	248	2 Q9L919	Q9L919 haemophilus
6	1205	89.8	263	2 Q9L918	Q9L918 haemophilus
7	225	16.8	295	2 Q8RR69	Q8RR69 actinobacil
8	225	16.8	295	2 Q8RO63	Q8RO63 actinobacil
9	222	16.5	504	2 Q9LA56	Q9LA56 escherichia
10	222	16.5	511	9 Q9MC18	Q9MC18 bacterioph
11	220	16.4	392	2 Q9LA60	Q9LA60 escherichia
12	217	16.2	355	2 Q8KUJ1	Q8KUJ1 haemophilus
13	217	16.2	362	2 Q8KH7	Q8KH7 neisseria m
14	217	16.2	364	16 Q9JXK7	Q9JXK7 neisseria m
15	217	16.2	459	2 Q8VW24	Q8VW24 escherichia
16	209	15.6	391	2 Q8KUJ6	Q8KUJ6 neisseria m

17	209	15.6	398	2	Q8KI42	Q8KI42 neisseria m
18	208.5	15.5	487	2	Q9LA53	Q9LA53 escherichia
19	204	15.2	405	2	Q8KHP5	Q8KHP5 neisseria m
20	204	15.2	405	2	Q8KH85	Q8KH85 neisseria m
21	198	14.8	668	2	Q8GH86	Q8GH86 moraxella c
22	196	14.6	630	2	Q9XD55	Q9XD55 moraxella c
23	193	14.4	613	2	Q9XD53	Q9XD53 moraxella c
24	193	14.4	616	2	Q848S2	Q848S2 moraxella c
25	193	14.4	877	2	Q848S1	Q848S1 moraxella c
26	193	14.4	889	2	Q9L961	Q9L961 moraxella c
27	193	14.4	894	2	Q9L962	Q9L962 moraxella c
28	190.5	14.2	576	2	Q54407	Q54407 moraxella c
29	190.5	14.2	674	2	Q9XD51	Q9XD51 moraxella c
30	190.5	14.2	684	2	Q9L963	Q9L963 moraxella c
31	190.5	14.2	686	2	Q8RTB2	Q8RTB2 moraxella c
32	186	13.9	422	2	Q93KR4	Q93KR4 moraxella c
33	186	13.9	422	2	Q56930	Q56930 yersinia en
34	180.5	13.5	422	2	Q84GR6	Q84GR6 yersinia en
35	161	12.0	454	2	O85267	O85267 yersinia en
36	159.5	11.9	2712	16	Q9F3X5	Q9F3X5 pasteurella
37	132	9.8	155	16	Q8YJ81	Q8YJ81 brucella me
38	130	9.7	278	16	Q8FYM1	Q8FYM1 brucella su
39	125.5	9.4	144	16	Q8CL85	Q8CL85 yersinia pe
40	124.5	9.3	1299	16	Q9F3X6	Q9F3X6 pasteurella
41	119.5	8.9	641	16	Q8CKM1	Q8CKM1 yersinia pe
42	119.5	8.9	658	16	Q8ZHJ0	Q8ZHJ0 yersinia pe
43	116	8.6	516	2	Q93F81	Q93F81 escherichia
44	112.5	8.4	364	16	Q8ZGB6	Q8ZGB6 yersinia pe
45	110	8.2	873	2	Q9L960	Q9L960 moraxella c

ALIGNMENTS

RESULT 1

Q9K2H6 PRELIMINARY; PRT; 257 AA.
ID Q9K2H6
AC Q9K2H6
DT 01-OCT-2000 (TRENBLrel. 15, Created)
DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Serum resistance protein DsrA (Serum resistance protein DsrA).
GN DSR A OR HD0769.
OS Haemophilus ducreyi.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=730;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CIP542(CDC), 35000, and 406;
RX MEDLINE=20143779; PubMed=10678980;
RA Elkins C., Morrow K.J. Jr., Olsen B.;
RT "Serum Resistance in Haemophilus ducreyi Requires Outer Membrane Protein DsrA."
RL Infect. Immun. 68:1608-1619(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=35000HP / ATCC 700724;
RA Munson R.S. Jr., Ray W.C., Mahairas G., Sabo P., Mungur R.,
RA Johnson L., Nguyen D., Wang J., Forst C., Hood L.;
RT "The complete genome sequence of Haemophilus ducreyi.";
RL Submitted (JUN-2003) to the EMBL/GenBank/DSJ databases.
DR EMBL; AF187007; AAF37813.1; -
DR EMBL; AF187001; AAF37807.1; -
DR EMBL; AF187005; AAF37811.1; -
DR EMBL; AE017152; AAP95674.1; -
DR InterPro; IPR005594; Yada.
DR Pfam; PF03895; Yada; 1.
KW Complete proteome.
SQ SEQUENCE 257 AA; 28524 MW; F497BF2CD5666938 CRC64;

Query Match 100.0%; Score 1342; DB 16; Length 257;
Best Local Similarity 100.0%; Pred. No. 5.4e-95;

```
Matches 257; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MKIKCLVAVGLACSTTTTMAQQPKFAGVSSLSYSEYDYGKWTWSNEGFDIKVPGI 60
Db 1 MKIKCLVAVGLACSTTTTMAQQPKFAGVSSLSYSEYDYGKWTWSNEGFDIKVPGI 60
Qy 61 KMKPKWISKQATYLELQHYMPYTPVLVITYAPGVSPSPILLYPMSDDPDLGINRQOLKLN 120
Db 61 KMKPKWISKQATYLELQHYMPYTPVLVITYAPGVSPSPILLYPMSDDPDLGINRQOLKLN 120
Qy 121 LYSYFNDLRHDFKLVLDARISKKNQIDTTSKYLLLELGTLYLDDSYRMEQNTNHNKNT 180
Db 121 LYSYFNDLRHDFKLVLDARISKKNQIDTTSKYLLLELGTLYLDDSYRMEQNTNHNKNT 180
Qy 181 KELOTGLANQSALSMVQPNVGVGKTSVSAAGVGYRDKTALAIGVGSRTIDRFTAKAGVAF 240
Db 181 KELOTGLANQSALSMVQPNVGVGKTSVSAAGVGYRDKTALAIGVGSRTIDRFTAKAGVAF 240
Qy 241 NTYNGGMSYGASVGVEF 257
Db 241 NTYNGGMSYGASVGVEF 257

RESULT 2
Q9K3B8
ID Q9K3B8 PRELIMINARY; PRT; 264 AA.
AC Q9K3B8;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Serum resistance protein Dera.
GN DERA.
OS Haemophilus ducreyi.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=730;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CIPA77; PubMed=10678980;
RX MEDLINE=20143779; PubMed=10678980;
RA Elkins C., Morrow K.J. Jr., Olsen B.;
RT "Serum Resistance in Haemophilus ducreyi Requires Outer Membrane Protein Dera.";
RL Infect. Immun. 68:1608-1619(2000).
DR EMBL; AF187003; AAF37809.1; -.
DR EMBL; AF187002; AAF37808.1; -.
DR InterPro; IPR005594; Yada.
DR Pfam; PF03895; Yada; 1.
SQ SEQUENCE 264 AA; 29444 MW; 11AF6C124D7ED126 CRC64;

Query Match 96.8%; Score 1299.5; DB 2; Length 264;
Best Local Similarity 95.8%; Pred. No. 1e-91;
Matches 253; Conservative 0; Mismatches 4; Indels 7; Gaps 1;
Qy 1 MKIKCLVAVGLACSTTTTMAQQPKFAGVSSLSYSEYDYGKWTWSNEGFDIKVPGI 60
Db 1 MKIKCLVAVGLACSTTTTMAQQPKFAGVSSLSYSEYDYGKWTWSNEGFDIKVPGI 60
Qy 61 KMKPKWISKQATYLELQHYMPYTPVLVITYAPGVSPSPILLYPMSDDPDLGINRQOLKLN 120
Db 61 KMKPKWISKQATYLELQHYMPYTPVLVITYAPGVSPSPILLYPMSDDPDLGINRQOLKLN 120
Qy 121 LYSYFNDLRHDFKLVLDARISKKNQIDTTSKYLLLELGTLYLDDSYRMEQNTNHNKNT 173
Db 121 LYSYFNDLRHDFKLVLDARISKKNQIDTTSKYLLLELGTLYLDDSYRMEQNTNHNKNT 180
Qy 174 HNINKLSKELQTLGLANQSALSMVQPNVGVGKTSVSAAGVGYRDKTALAIGVGSRTIDRFT 233
Db 181 HNINKLSKELQTLGLANQSALSMVQPNVGVGKTSVSAAGVGYRDKTALAIGVGSRTIDRFT 240
Qy 234 AKAGVAFNTYNGGMSYGASVGVEF 257
Db 241 AKAGVAFNTYNGGMSYGASVGVEF 264
```

RESULT 3

```
Q9L916
ID Q9L916 PRELIMINARY; PRT; 271 AA.
AC Q9L916;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Serum resistance protein Dera.
GN DERA.
OS Haemophilus ducreyi.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=730;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=V1157;
RX MEDLINE=20143779; PubMed=10678980;
RA Elkins C., Morrow K.J. Jr., Olsen B.;
RT "Serum Resistance in Haemophilus ducreyi Requires Outer Membrane Protein Dera.";
RL Infect. Immun. 68:1608-1619(2000).
DR EMBL; AF187009; AAF37815.1; -.
DR InterPro; IPR005594; Yada.
DR Pfam; PF03895; Yada; 1.
SQ SEQUENCE 271 AA; 30091 MW; 5A9DC55C6DA6EF1C CRC64;
```

Query Match 96.1%; Score 1289; DB 2; Length 271;
Best Local Similarity 93.0%; Pred. No. 6.7e-91;
Matches 252; Conservative 0; Mismatches 5; Indels 14; Gaps 1;

```
Qy 1 MKIKCLVAVGLACSTTTTMAQQPKFAGVSSLSYSEYDYGKWTWSNEGFDIKVPGI 60
Db 1 MKIKCLVAVGLACSTTTTMAQQPKFAGVSSLSYSEYDYGKWTWSNEGFDIKVPGI 60
Qy 61 KMKPKWISKQATYLELQHYMPYTPVLVITYAPGVSPSPILLYPMSDDPDLGINRQOLKLN 120
Db 61 KMKPKWISKQATYLELQHYMPYTPVLVITYAPGVSPSPILLYPMSDDPDLGINRQOLKLN 120
Qy 121 LYSYFNDLRHDFKLVLDARISKKNQIDTTSKYLLLELGTLYLDDSYRMEQNTNHNKNT 171
Db 121 LYSYFNDLRHDFKLVLDARISKKNQIDTTSKYLLLELGTLYLDDSYRMEQNTNHNKNT 180
Qy 172 -----NTNHNKLSKELQTLGLANQSALSMVQPNVGVGKTSVSAAGVGYRDKTALAIGVGS 226
Db 181 HNINKNTNHNKLSKELQTLGLANQSALSMVQPNVGVGKTSVSAAGVGYRDKTALAIGVGS 240
Qy 227 RITDRFTAKGVAFNTYNGGMSYGASVGVEF 257
Db 241 RITDRFTAKGVAFNTYNGGMSYGASVGVEF 271
```

RESULT 4

```
Q9L917
ID Q9L917 PRELIMINARY; PRT; 273 AA.
AC Q9L917;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Serum resistance protein Dera.
GN DERA.
OS Haemophilus ducreyi.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=730;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=M90-02;
RX MEDLINE=20143779; PubMed=10678980;
RA Elkins C., Morrow K.J. Jr., Olsen B.;
RT "Serum Resistance in Haemophilus ducreyi Requires Outer Membrane Protein Dera.";
```



```
RL Infect. Immun. 68:1608-1619 (2000).
DR EMBL: AF187008; AAF37814.1; -.
DR InterPro: IPR005594; Yada.
DR Pfam: PF03895; Yada; 1.
SQ SEQUENCE 273 AA; 30281 MW; 8C620D6B5544E3EE CRC64;

Query Match 95.6%; Score 1283; DB 2; Length 273;
Best Local Similarity 92.7%; Pred. No. 1.9e-90;
Matches 253; Conservative 0; Mismatches 4; Indels 16; Gaps 2;

Qy 1 MKIKLVAVVGLACSTTTTMAQQPPKFAVSSLSYSEYDYGKGTWWSNCGGFDIKVPGI 60
Db 1 MKIKLVAVVGLACSTTTTMAQQPPKFAVSSLSYSEYDYGKGTWWSNCGGFDIKVPGI 60
Qy 61 KMKPKWISKQATYLELQHYMPYTPVLVTPAGVSPS--PLLVPMSDDPOLGIRNROOLK 118
Db 61 KMKPKWISKQATYLELQHYMPYTPVLVTPAGVSPS--PLLVPMSDDPOLGIRNROOLK 120
Qy 119 LNLVSFNDLRHDFKLVLDARISKNQNDITISKYLLELGTLYLDDSYRMMEQ----- 171
Db 121 LNLVSFNDLRHDFKLVLDARISKNQNDITISKYLLELGTLYLDDSYRMMEQNTHNINK 180
Qy 172 -----NTNINKLSKELQTLGLANQSALSMVQPNVGKTSVSAAVGGYRDKTALAIGV 224
Db 181 NTNINKNTNINKLSKELQTLGLANQSALSMVQPNVGKTSVSAAVGGYRDKTALAIGV 240
Qy 225 GSRITDRFTAKAGVAFTNYNGMSYGASVGVEF 257
Db 241 GSRITDRFTAKAGVAFTNYNGMSYGASVGVEF 273

RESULT 5
Q9L919 PRELIMINARY; PRT; 248 AA.
ID Q9L919;
AC Q9L919;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE Serum resistance protein DsrA.
GN DsrA.
OS Haemophilus ducreyi.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=730;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=C14;
RX MEDLINE=20143779; PubMed=10678980;
RA Elkins C., Morrow K.J. Jr., Olsen B.;
RT "Serum Resistance in Haemophilus ducreyi Requires Outer Membrane Protein DsrA.";
RL Infect. Immun. 68:1608-1619 (2000).
DR EMBL: AF187008; AAF37814.1; -.
DR InterPro: IPR005594; Yada.
DR Pfam: PF03895; Yada; 1.
SQ SEQUENCE 248 AA; 27815 MW; 245C809F2486A815 CRC64;

Query Match 94.0%; Score 1262; DB 2; Length 248;
Best Local Similarity 99.6%; Pred. No. 7e-89;
Matches 242; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MKIKLVAVVGLACSTTTTMAQQPPKFAVSSLSYSEYDYGKGTWWSNCGGFDIKVPGI 60
Db 1 MKIKLVAVVGLACSTTTTMAQQPPKFAVSSLSYSEYDYGKGTWWSNCGGFDIKVPGI 60
Qy 61 KMKPKWISKQATYLELQHYMPYTPVLVTPAGVSPSPLLVPMSDDPOLGIRNROOLK 120
Db 61 KMKPKWISKQATYLELQHYMPYTPVLVTPAGVSPSPLLVPMSDDPOLGIRNROOLK 120
Qy 121 LYSFNDLRHDFKLVLDARISKNQNDITISKYLLELGTLYLDDSYRMMEQNTHNINKLS 180
Db 121 LYSFNDLRHDFKLVLDARISKNQNDITISKYLLELGTLYLDDSYRMMEQNTHNINKLS 180

RL Infect. Immun. 68:1608-1619 (2000).
DR EMBL: AF187008; AAF37814.1; -.
DR InterPro: IPR005594; Yada.
DR Pfam: PF03895; Yada; 1.
SQ SEQUENCE 248 AA; 27815 MW; 245C809F2486A815 CRC64;

Query Match 94.0%; Score 1262; DB 2; Length 248;
Best Local Similarity 99.6%; Pred. No. 7e-89;
Matches 242; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MKIKLVAVVGLACSTTTTMAQQPPKFAVSSLSYSEYDYGKGTWWSNCGGFDIKVPGI 60
Db 1 MKIKLVAVVGLACSTTTTMAQQPPKFAVSSLSYSEYDYGKGTWWSNCGGFDIKVPGI 60
Qy 61 KMKPKWISKQATYLELQHYMPYTPVLVTPAGVSPSPLLVPMSDDPOLGIRNROOLK 120
Db 61 KMKPKWISKQATYLELQHYMPYTPVLVTPAGVSPSPLLVPMSDDPOLGIRNROOLK 120
Qy 121 LYSFNDLRHDFKLVLDARISKNQNDITISKYLLELGTLYLDDSYRMMEQNTHNINKLS 180
Db 121 LYSFNDLRHDFKLVLDARISKNQNDITISKYLLELGTLYLDDSYRMMEQNTHNINKLS 180

Infect. Immun. 68:1608-1619 (2000).
DR EMBL: AF187008; AAF37814.1; -.
DR InterPro: IPR005594; Yada.
DR Pfam: PF03895; Yada; 1.
SQ SEQUENCE 273 AA; 30281 MW; 8C620D6B5544E3EE CRC64;

Query Match 95.6%; Score 1283; DB 2; Length 273;
Best Local Similarity 92.7%; Pred. No. 1.9e-90;
Matches 253; Conservative 0; Mismatches 4; Indels 16; Gaps 2;

Qy 1 MKIKLVAVVGLACSTTTTMAQQPPKFAVSSLSYSEYDYGKGTWWSNCGGFDIKVPGI 60
Db 1 MKIKLVAVVGLACSTTTTMAQQPPKFAVSSLSYSEYDYGKGTWWSNCGGFDIKVPGI 60
Qy 61 KMKPKWISKQATYLELQHYMPYTPVLVTPAGVSPS--PLLVPMSDDPOLGIRNROOLK 118
Db 61 KMKPKWISKQATYLELQHYMPYTPVLVTPAGVSPS--PLLVPMSDDPOLGIRNROOLK 120
Qy 119 LNLVSFNDLRHDFKLVLDARISKNQNDITISKYLLELGTLYLDDSYRMMEQ----- 171
Db 121 LNLVSFNDLRHDFKLVLDARISKNQNDITISKYLLELGTLYLDDSYRMMEQNTHNINK 180
Qy 172 -----NTNINKLSKELQTLGLANQSALSMVQPNVGKTSVSAAVGGYRDKTALAIGV 224
Db 181 NTNINKNTNINKLSKELQTLGLANQSALSMVQPNVGKTSVSAAVGGYRDKTALAIGV 240
Qy 225 GSRITDRFTAKAGVAFTNYNGMSYGASVGVEF 257
Db 241 GSRITDRFTAKAGVAFTNYNGMSYGASVGVEF 273

RESULT 6
Q9L918 PRELIMINARY; PRT; 263 AA.
ID Q9L918;
AC Q9L918;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE Serum resistance protein DsrA.
GN DsrA.
OS Haemophilus ducreyi.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=730;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=CHIA;
RX MEDLINE=20143779; PubMed=10678980;
RA Elkins C., Morrow K.J. Jr., Olsen B.;
RT "Serum Resistance in Haemophilus ducreyi Requires Outer Membrane Protein DsrA.";
RL Infect. Immun. 68:1608-1619 (2000).
DR EMBL: AF187008; AAF37814.1; -.
DR InterPro: IPR005594; Yada.
DR Pfam: PF03895; Yada; 1.
SQ SEQUENCE 263 AA; 29170 MW; 858B91BF765B89F1 CRC64;

Query Match 89.8%; Score 1205; DB 2; Length 263;
Best Local Similarity 89.0%; Pred. No. 1.8e-84;
Matches 235; Conservative 8; Mismatches 13; Indels 8; Gaps 2;

Qy 1 MKIKLVAVVGLACSTTTTMAQQPPKFAVSSLSYSEYDYGKGTWWSNCGGFDIKVPGI 60
Db 1 MKIKLVAVVGLACSTTTTMAQQPPKFAVSSLSYSEYDYGKGTWWSNCGGFDIKVPGI 60
Qy 61 KMKPKWISKQATYLELQHYMPYTPVLVTPAGVSPSPLLVPMSDDPOLGIRNROOLK 120
Db 61 KMKPKWISKQATYLELQHYMPYTPVLVTPAGVSPSPLLVPMSDDPOLGIRNROOLK 119
Qy 121 LYSFNDLRHDFKLVLDARISKNQNDITISKYLLELGTLYLDDSYRMMEQ-----NT 173
Db 120 LYSFNDLRHDFKLVLDARISKNQNDITISKYLLELGTLYLDDSYRMMEQNTHNINK 179
Qy 174 HNINKLSKELQTLGLANQSALSMVQPNVGKTSVSAAVGGYRDKTALAIGVSRITDRFT 233
Db 180 HNINKLSKELQTLGLANQSALSMVQPNVGKTSVSAAVGGYRDKTALAIGVSRITDRFT 239
Qy 234 AKAGVAFNTYNGMSYGASVGVEF 257
Db 240 AKAGVAFNTYNGMSYGASVGVEF 263

RESULT 7
Q8RR69 PRELIMINARY; PRT; 295 AA.
ID Q8RR69;
AC Q8RR69;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DE Outer membrane protein 100.
GN OMP100.
OS Actinobacillus actinomycetemcomitans (Haemophilus actinomycetemcomitans).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Actinobacillus.
```

```
OX NCBI_TaxID=714;
RP SEQUENCE FROM N.A.
RA Komatsuzawa H., Asakawa R., Kawai T., Ochiai K., Fujiwara T.,
RT Taubman M.A., Kurihara H., Sugai M.;
RT "Identification of six major outer membrane proteins from
RL Actinobacillus actinomycetemcomitans.";
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB064943; BAB86905.1; -; Ubiq_surface.
DR InterPro; IPR008378; Ubiq_surface.
DR InterPro; IPR005594; YadaA.
DR Pfam; PF03895; YadaA; 1.
DR PRINTS; PR01804; UBIQUITOUSSP.
SQ SEQUENCE 295 AA; 32062 MW; C9AB93C40D7E1C69 CRC64;

Query Match 16.8%; Score 225; DB 2; Length 295;
Best Local Similarity 31.8%; Pred. No. 3e-09;
Matches 55; Conservative 34; Mismatches 54; Indels 30; Gaps 2;

Qy 111 GINRQQLNLNLSYFNDLRHDPKLVLDARISKKNQKID----- 149
Db 127 GVNRLQNVDRSTENAARS-----RANEQKIAENKKALENKADKADVEKNRADIANSRA 182

Qy 150 -----TISKYLLELGTYLDDSYRMEQNTNINKLSKELOTLQGLANQSALSMVLPNGYVK 204
Db 183 IATFRSSQNTIAALTTKVDRTNARTIDRLDSRVNLDKEVKNGLASQAALSGLFQPNYVGS 242

Qy 205 TSVSAAGVGYRDKTALAIGVGSRTIDRTAKAGVAFNTYNGMSYGASVGYEF 257
Db 243 LNLSSAAGVGYKSTALAVGSGYRPNQNVAAKAGVAVSTNGGSATYVGLNFEW 295

RESULT 8
Qy Q8RQ63 PRELIMINARY; PRT; 295 AA.
AC Q8RQ63;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Putative adhesin/invasin.
GN AA32-1-2.
OS Actinobacillus actinomycetemcomitans (Haemophilus
OC actinomycetemcomitans).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Actinobacillus.
OX NCBI_TaxID=714;
RP SEQUENCE FROM N.A.
RA Lepine G., Li L., Ellen R.P.;
RT "Cloning and characterization of three invasive genes of
RL Actinobacillus actinomycetemcomitans.";
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF316502; AAL91673.1; -.
DR InterPro; IPR008378; Ubiq_surface.
DR InterPro; IPR005594; YadaA.
DR Pfam; PF03895; YadaA; 1.
DR PRINTS; PR01804; UBIQUITOUSSP.
SQ SEQUENCE 295 AA; 32034 MW; C9875DD9DD7CCDBA CRC64;

Query Match 16.8%; Score 225; DB 2; Length 295;
Best Local Similarity 31.8%; Pred. No. 3e-09;
Matches 55; Conservative 34; Mismatches 54; Indels 30; Gaps 2;

Qy 111 GINRQQLNLNLSYFNDLRHDPKLVLDARISKKNQKID----- 149
Db 127 GVNRLQNVDRSTENAARS-----RANEQKIAENKKALENKADKADVEKNRADIANSRA 182

Qy 150 -----TISKYLLELGTYLDDSYRMEQNTNINKLSKELOTLQGLANQSALSMVLPNGYVK 204
Db 183 IATFRSSQNTIAALTTKVDRTNARTIDRLDSRVNLDKEVKNGLASQAALSGLFQPNYVGS 242

Qy 205 TSVSAAGVGYRDKTALAIGVGSRTIDRTAKAGVAFNTYNGMSYGASVGYEF 257
```

```
Db 243 LNLSSAAGVGYKSTALAVGSGYRPNQNVAAKAGVAVSTNGGSATYVGLNFEW 295

RESULT 9
Qy Q9LA56 PRELIMINARY; PRT; 504 AA.
AC Q9LA56;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Immunoglobulin-binding protein EibC.
GN EIBC.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RP SEQUENCE FROM N.A.
RA STRAIN=ECOR-9;
RC MEDLINE=20187497; PubMed=10722621;
RA Sandt C.H., Hill C.W.;
RT "Four different genes responsible for nonimmune immunoglobulin-binding
RT activities within a single strain of Escherichia coli.";
RL Infect. Immun. 68:2205-2214 (2000).
DR EMBL; AF151674; AAF63035.1; -.
DR InterPro; IPR008640; Hep_Hag.
DR InterPro; IPR005594; YadaA.
DR Pfam; PF05658; Hep_Hag; 4.
DR Pfam; PF03895; YadaA; 1.
SQ SEQUENCE 504 AA; 53158 MW; 6B5E192DD4B2771B CRC64;

Query Match 16.5%; Score 222; DB 2; Length 504;
Best Local Similarity 38.0%; Pred. No. 1e-08;
Matches 49; Conservative 26; Mismatches 52; Indels 2; Gaps 1;

Qy 129 RHDFKLKVLDAISKKNQKIDITISKYLLELGTLYLDDSYRMEQNTNINKLSKELOTLGLA 188
Db 378 RIDYAVGVIDGRVTRNTQSIKSKAKAANTRTIQHSARLDSQQRQINENHKEMKRAA 437

Qy 189 NQSALSMVLPNGVGVKTSVSAAGVGYRDKTALAIGVGSRTIDRTAKAGVAFNTYNGMS 248
Db 438 QSAALTGLFQPYSGVGFNATAAGVGSYDQQAALAVGVGYRFEQTAAGAGVAFS--DGDAS 495

Qy 249 YGASVGYEF 257
Db 496 WNVGVNFEF 504

RESULT 10
Qy Q9MC18 PRELIMINARY; PRT; 511 AA.
AC Q9MC18;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Immunoglobulin-binding protein EibD.
GN EIBD.
OS Bacteriophage P-EibD.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae.
OX NCBI_TaxID=120163;
RP SEQUENCE FROM N.A.
RA MEDLINE=20187497; PubMed=10722621;
RA Sandt C.H., Hill C.W.;
RT "Four different genes responsible for nonimmune immunoglobulin-binding
RT activities within a single strain of Escherichia coli.";
RL Infect. Immun. 68:2205-2214 (2000).
DR EMBL; AF151675; AAF63040.1; -.
DR InterPro; IPR008640; Hep_Hag.
DR InterPro; IPR005594; YadaA.
DR Pfam; PF05658; Hep_Hag; 3.
DR Pfam; PF03895; YadaA; 1.
```

[illegible][illegible]

```
Qy 187 LANQSALSMVQPNVGKTSYSAAGGYRDKTALAIGVGSRTDRFTAKAGVAFNTYNG- 245
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 291 LAEQAALSGLFQPNVNGRFNFTAAVGGYKSESAVAIGTGFRFTENFAAKAGVAVGTSSGS 350

Qy 246 GMSYGASVGYEF 257
:|||||:
Db 351 SAAHYHGVGVYEW 362

RESULT 14
Q9JXK7 PRELIMINARY; PRT; 364 AA.
AC Q9JXK7;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Adhesin/invasin, putative.
GN NMB1994.
OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC Neisseriaceae; Neisseria.
OX NCBI_TaxID=491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MC58 / Serogroup B;
RX MEDLINE=2017555; PubMed=10710307;
RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
RA Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
RA Nelson W.C., Gwinn M.L., DeBoy R., Peterson J.D., Hickey E.K.,
RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA Mason T., Ciecko A., Parksey D.S., Blair E., Citterone H., Clark E.B.,
RA Coston M.D., Utterback T.R., Khouri H., Qin H., Vamathevan J.,
RA Gill J., Scarlato V., Maignani V., Fizza M., Grandi G., Sun L.,
RA Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;
RT "Complete genome sequence of Neisseria meningitidis serogroup B strain
RT MC58.";
RL Science 287:1809-1815 (2000).
DR EMBL; AE002548; AAP42321.1; -.
DR FIR; AB1019; AB1019.
DR TIGR; NMB1994; -.
DR InterPro; IPR005594; Yada.
DR Pfam; PF03895; Yada; 1.
KW Complete proteome.
SQ SEQUENCE 364 AA; 37972 MW; B9FF591FD25F853F CRC64;

Query Match 16.2%; Score 217; DB 16; Length 364;
Best Local Similarity 39.4%; Pred. NO. 1.6e-08;
Matches 52; Conservative 19; Mismatches 37; Indels 24; Gaps 3;

Qy 127 DLRHDFKLKVLDAISKKNQIDTISKYLLLELGTLYLDDSYRMMQNTNINKLSKELOTG 186
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 256 DIKADIATNKAD--IAKNSARIDSLDK-----NVAHLRKETROG 292

Qy 187 LANQSALSMVQPNVGKTSYSAAGGYRDKTALAIGVGSRTDRFTAKAGVAFNTYNG- 245
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 293 LAEQAALSGLFQPNVNGRFNFTAAVGGYKSESAVAIGTGFRFTENFAAKAGVAVGTSSGS 352

Qy 246 GMSYGASVGYEF 257
:|||||:
Db 353 SAAHYHGVGVYEW 364

RESULT 15
Q8VW24 PRELIMINARY; PRT; 459 AA.
AC Q8VW24;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Immunoglobulin-binding protein Eibf.
GN EIBF.
OS Escherichia coli.
```

```
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ECOR-2;
RX MEDLINE=21562582; PubMed=11705900;
RA Sandt C.H., Hill C.W.;
RT "Nonimmune Binding of Human IgA and IGG Fc by Distinct Sequence
RT Segments of the Eibf Cell Surface Protein of Escherichia coli.";
RL Infect. Immun. 69:7293-7303 (2001).
DR EMBL; AF399847; AAL31493.1; -.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0001584; F:rhodopsin-like receptor activity; IEA.
DR GO; GO:0007186; P:G-protein coupled receptor protein signalin. .; IEA.
DR InterPro; IPR000276; GPCR_Rhodpsn.
DR InterPro; IPR005594; Yada.
DR Pfam; PF03895; Yada; 1.
DR PROSITE; PS00237; G_PROTEIN_RECEP_F1_1; 1.
SQ SEQUENCE 459 AA; 49489 MW; C149D12B564D19BB CRC64;

Query Match 16.2%; Score 217; DB 2; Length 459;
Best Local Similarity 38.0%; Pred. No. 2.2e-08;
Matches 49; Conservative 25; Mismatches 53; Indels 2; Gaps 1;

Qy 129 RHDFKLKVLDAISKKNQIDTISKYLLLELGTLYLDDSYRMMQNTNINKLSKELOTGLA 188
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 333 RIDAAGVAGIDGRVTRNTQSTIEKSKAIAANTRTLQOHSARLDSOQRQINENHKEMKRAAA 392

Qy 189 NQSALSMVQPNVGKTSVSAAGGYRDKTALAIGVGSRTDRFTAKAGVAFNTYNGMS 248
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 393 QSAALTGLFQFYSYVKFNATAAAGGYSDDQALAVGVGVYRFEQTAAKAGVAFS--DGDAS 450

Qy 249 YGASVGYEF 257
:|||||:
Db 451 WNVGVNPEF 459

Search completed: May 13, 2004, 08:17:28
Job time : 47 secs
```



```
Alignment Scores:
Pred. No.: 2.72e-11 Length: 610
Score: 193.00 Matches: 53
Percent Similarity: 50.71% Conservative: 18
Best Local Similarity: 37.86% Mismatches: 61
Query Match: 9.42% Indels: 8
DB: 10 Gaps: 3

US-10-030-529A-1 (1-1168) x US-09-952-267-11 (1-610)

Qy 476 AACGATTTAAGACACGAGTTTAAATTAATAAGTTCTTGATGCACGATTTTCCAAAAATAAA 535
Db 485 AsnLysAlaSerAlaAspThrLysPheAlaAlaThrAlaAspAlaIleThrLysAsnGly 504
Qy 536 CAAAATATTGATCTATAAGTAAATATTACTAGAACTGGGTACTTATTTAGATGAT--- 592
Db 505 AsnAlaIleThrLysAsnAlaLysSerIleThrAspLeuGlyThrLysValAspGlyPhe 524
Qy 593 TCTTATCGTATGATGGACAAATAACACAT-----AATATCAATAAG 634
Db 525 AspGlyArgValThrAlaLeuAspThrLysValAsnAlaPheAspGlyArgIleThrAla 544
Qy 635 TTGTCTAAAGAAATTCGCAAACTGGTTTAGCCCAACCAATCAGCATTGCTGTATGTTAGTCAA 694
Db 545 LeuAspSerLysValGluAsnGlyMetAlaAlaGlnAlaLeuSerGlyLeuPheGln 564
Qy 695 CAAAATGGTGTAGCAAAACGAGCGTTTCTGCTGCGGTAGGAGTTATAGAGATAAAACT 754
Db 565 ProTyrSerValGlyLysPheAsnAlaThrAlaAlaLeuGlyGlyTyrGlySerLysSer 584
Qy 755 GCATTAGCCATTGTGTCGCTCACGCATTACTGATCGCTTTACCGCTAAAGCGGGTGA 814
Db 585 AlaValAlaIleGlyAlaGlyTyrArgValAsnProAsnLeuAlaPheLysAlaGlyAla 604
Qy 815 GCGTTCAATACCTAC---AATGGCGGCATGCTTCTTATGCTGCTCTGTGGTTATGAATTC 871
Db 605 AlaIleAsnThrSerGlyAsnLysLysGlySerTyrAsnIleGlyValAsnTyrGluPhe 624

RESULT 3
US-09-952-267-15
; Sequence 15, Application US/09952267
; Publication No. US20030032772A1
; GENERAL INFORMATION:
; APPLICANT: HANSEN, ERIC J.
; APPLICANT: ABEI, CHRISTOPH
; APPLICANT: COPE, LESLIE D.
; APPLICANT: MACIVER, ISOBEL
; APPLICANT: FISKE, MICHAEL J.
; APPLICANT: FREDENBURG, ROSS A.
; TITLE OF INVENTION: USP1 AND USP2 ANTIGENS OF MORAXELLA CATARRHALIS
; FILE REFERENCE: AMCY:024
; CURRENT APPLICATION NUMBER: US/09/952,267
; CURRENT FILING DATE: 2001-09-12
; PRIOR APPLICATION NUMBER: 09/336,447
; PRIOR FILING DATE: 1999-06-21
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 15
; LENGTH: 889
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-09-952-267-15

Alignment Scores:
Pred. No.: 3.01e-11 Length: 889
Score: 193.00 Matches: 53
Percent Similarity: 50.71% Conservative: 18
Best Local Similarity: 37.86% Mismatches: 61
Query Match: 9.42% Indels: 8
DB: 10 Gaps: 3

US-10-030-529A-1 (1-1168) x US-09-952-267-15 (1-889)

Qy 476 AACGATTTAAGACACGAGTTTAAATTAATAAGTTCTTGATGCACGATTTTCCAAAAATAAA 535
Db 750 AsnLysAlaSerAlaAspThrLysPheAlaAlaThrAlaAspAlaIleThrLysAsnGly 769
Qy 536 CAAAATATTGATCTATAAGTAAATATTACTAGAACTGGGTACTTATTTAGATGAT--- 592
Db 770 AsnAlaIleThrLysAsnAlaLysSerIleThrAspLeuGlyThrLysValAspGlyPhe 789
Qy 593 TCTTATCGTATGATGGACAAATAACACAT-----AATATCAATAAG 634

Alignment Scores:
Pred. No.: 2.74e-11 Length: 624
Score: 193.00 Matches: 53
Percent Similarity: 50.71% Conservative: 18
Best Local Similarity: 37.86% Mismatches: 61
Query Match: 9.42% Indels: 8
DB: 10 Gaps: 3

US-10-030-529A-1 (1-1168) x US-09-952-267-7 (1-624)

Qy 476 AACGATTTAAGACACGAGTTTAAATTAATAAGTTCTTGATGCACGATTTTCCAAAAATAAA 535
Db 471 AsnLysAlaSerAlaAspThrLysPheAlaAlaThrAlaAspAlaIleThrLysAsnGly 490
Qy 536 CAAAATATTGATCTATAAGTAAATATTACTAGAACTGGGTACTTATTTAGATGAT--- 592
Db 491 AsnAlaIleThrLysAsnAlaLysSerIleThrAspLeuGlyThrLysValAspGlyPhe 510
Qy 593 TCTTATCGTATGATGGACAAATAACACAT-----AATATCAATAAG 634
Db 511 AspGlyArgValThrAlaLeuAspThrLysValAsnAlaPheAspGlyArgIleThrAla 530
Qy 635 TTGTCTAAAGAAATTCGCAAACTGGTTTAGCCCAACCAATCAGCATTGCTGTATGTTAGTCAA 694
Db 531 LeuAspSerLysValGluAsnGlyMetAlaAlaGlnAlaLeuSerGlyLeuPheGln 550
Qy 695 CAAAATGGTGTAGCAAAACGAGCGTTTCTGCTGCGGTAGGAGTTATAGAGATAAAACT 754
Db 551 ProTyrSerValGlyLysPheAsnAlaThrAlaAlaLeuGlyGlyTyrGlySerLysSer 570
Qy 755 GCATTAGCCATTGTGTCGCTCACGCATTACTGATCGCTTTACCGCTAAAGCGGGTGA 814
Db 571 AlaValAlaIleGlyAlaGlyTyrArgValAsnProAsnLeuAlaPheLysAlaGlyAla 590
Qy 815 GCGTTCAATACCTAC---AATGGCGGCATGCTTATGCTGCTCTGTGGTTATGAATTC 871
Db 591 AlaIleAsnThrSerGlyAsnLysLysGlySerTyrAsnIleGlyValAsnTyrGluPhe 610

RESULT 2
US-09-952-267-7
; Sequence 7, Application US/09952267
; Publication No. US20030032772A1
; GENERAL INFORMATION:
; APPLICANT: HANSEN, ERIC J.
; APPLICANT: ABEI, CHRISTOPH
; APPLICANT: COPE, LESLIE D.
; APPLICANT: MACIVER, ISOBEL
; APPLICANT: FISKE, MICHAEL J.
; APPLICANT: FREDENBURG, ROSS A.
; TITLE OF INVENTION: USP1 AND USP2 ANTIGENS OF MORAXELLA CATARRHALIS
; FILE REFERENCE: AMCY:024
; CURRENT APPLICATION NUMBER: US/09/952,267
; CURRENT FILING DATE: 2001-09-12
; PRIOR APPLICATION NUMBER: 09/336,447
; PRIOR FILING DATE: 1999-06-21
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
; LENGTH: 624
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-09-952-267-7
```



```

QY 422 GATCACTTTGGAATAAT-----CGGCAGCAGCTGAAATTTGAAT 460
Db 2564 AspValLysGlyLeuValThrGluAsnLeuThrValLysGluGlyAlaLysIleAsn 2583
QY 461 TTGTATAGTTATTTAAACGAT-----TTAAGACACGATTTTAAATA 502
Db 2584 MetGlyAsnAsnValIleAspGlyValAlaAspGlyGluValAsnAlaThrSerLysGln 2603
QY 503 AAAGTTCTTGATCGAGTATTTCCAAAATAAACAATAATATTGATCTATTAAGTAAATAT 562
Db 2604 AlaValAsnGlySerGlnLeuHisLysValGlnGlnGlnValAsnGlnAlaThrAla 2623
QY 563 TTACTAGAACTGGTACTTTATTAGATGATTTCTTATCGTATGATGAAACAAAATACACAT 622
Db 2624 IleAsnLysLeuGlyAspHis-----2630
QY 623 AATATCAATAGTTGCTTAAAGAAATGCAACTGTTTAGCCCAACCAATCAGCATTTGCT 682
Db 2631 ---IleAsnLysValAspLysAspLeuArgAlaGlyIleAlaGlyAlaThrAlaValAla 2649
QY 683 ATGTTAGTCAACCAATCGTGTAGGCAAAACGAGCGTTTCTGCTGCGGTAGGAGTTAT 742
Db 2650 PheLeuGlnArgProGlnGluAlaGlyLysSerIleValSerLeuGlyValGlySerTyr 2669
QY 743 AGAGATAAAACTGCAATTAGCCATTGTTGTCGCTCAGCATTTACTGAT-----CGCTTT 796
Db 2670 ArgSerGluSerAlaIleAlaValGlyTyr---AlaArgAsnSerAspAsnAsnLysIle 2688
QY 797 ACGGTAAGCGGGTGTAGCGTTCAATACCTACATGCGCGCATGCTTTATGCTGCTTCT 856
Db 2689 SerIleLysLeuGlyGlyGlyMetAsnSer---ArgGlyAspValAsnPheGlyGlySer 2707
QY 857 GTTGGTTATGAATTC 871
Db 2708 IleGlyTyrGlnTrp 2712

```

RESULT 6

```

US-10-192-584-6
; Sequence 6, Application US/10192584
; Publication No. US20030027987A1
; GENERAL INFORMATION:
; APPLICANT: TOKUNAGA, Eiichi
; SAKAGUCHI, Masashi
; MATSUO, Kazuo
; HAMADA, Fukusaburo
; TOKIYOSHI, Sachio
; TITLE OF INVENTION: NOVEL POLYPEPTIDE FROM HAEMOPHILUS
; PARAGALLINARUM AND PROCESS FOR PREPARING THE SAME
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 624 Ninth Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/192,584
; FILING DATE: 11-Jul-2002
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/077,098
; FILING DATE: 19-May-1998
; APPLICATION NUMBER: PCT/JP97/03222
; FILING DATE: 12-SEP-1997
; APPLICATION NUMBER: JP 27,148/1996
; FILING DATE: 19-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: KORNBAU, Anne M.

```

```

; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: TOKUNAGA=1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; TYPE: amino acid
; LENGTH: 2042 amino acids
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-10-192-584-6
Alignment Scores:
Pred. No.: 0.00702 Length: 2042
Score: 115.00 Matches: 31
Percent Similarity: 49.04% Conservative: 20
Best Local Similarity: 29.81% Mismatches: 47
Query Match: 5.62% Indels: 6
DB: 14 Gaps: 3
US-10-030-529A-1 (1-1168) x US-10-192-584-6 (1-2042)
QY 569 GAACCTGGTACTTATTAGATGATTTCTTATCGTATGATGGAACAAATACACATAATATC 628
Db 1942 GlnLeuAsnThrValIleAspAsnValGlnAsnAsnPheAsnGlnValAsnGlnArgIle 1961
QY 629 AATAAGTTGCTTAAAGAAATGCAACTGTTTAGCCCAACCAATCAGCATTTCTTATGTTA 688
Db 1962 GlyAspLeuThrArgGluSerArgAlaGlyIleAlaGlyAlaMetAlaThrAlaSerLeu 1981
QY 689 GTGCAACCAATGCTGTAGGCAAAACGAGCGTTTCTGCTGCGGTAGGAGTTATAGAGAT 748
Db 1982 GlnAsnValAlaLeuProGlyLysThrIleSerValGlyThrAlaThrPheLysGly 2001
QY 749 AAAAAGCTGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 808
Db 2002 GluAsnAlaValAlaIleGlyMet---SerArgLeuSerAsp-----AsnGlyLysVal 2018
QY 809 GGTGTA-----GGCTTCAATACCTACATGCGCGCATGCTTATGGTGGTCTTCTGTT 859
Db 2019 GlyIleArgLeuSerGlyMetSerThrSerAsnGlyAspLysGlyAlaAlaMetSerVal 2038
QY 860 GGTATGAAATTC 871
Db 2039 GlyPheSerPhe 2042
RESULT 7
US-10-192-584-7
; Sequence 7, Application US/10192584
; Publication No. US20030027987A1
; GENERAL INFORMATION:
; APPLICANT: TOKUNAGA, Eiichi
; SAKAGUCHI, Masashi
; MATSUO, Kazuo
; HAMADA, Fukusaburo
; TOKIYOSHI, Sachio
; TITLE OF INVENTION: NOVEL POLYPEPTIDE FROM HAEMOPHILUS
; PARAGALLINARUM AND PROCESS FOR PREPARING THE SAME
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 624 Ninth Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/192,584
; FILING DATE: 11-Jul-2002
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/077,098
; FILING DATE: 19-May-1998
; APPLICATION NUMBER: PCT/JP97/03222
; FILING DATE: 12-SEP-1997
; APPLICATION NUMBER: JP 27,148/1996
; FILING DATE: 19-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: KORNBAU, Anne M.

```


SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA: US/10/192,584
FILING DATE: 11-Jul-2002
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/077,098
FILING DATE: 19-May-1998
APPLICATION NUMBER: PCT/Jp97/03222
FILING DATE: 12-SEP-1997
APPLICATION NUMBER: JP 27,148/1996
FILING DATE: 19-SEP-1996
ATTORNEY/AGENT INFORMATION:
NAME: KORNEAU, Anne M.
REGISTRATION NUMBER: 25,618
REFERENCE/DOCKET NUMBER: TOKUNAGA-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 2039 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 7:
US-10-192-584-7

Alignment Scores:
Pred. No.: 0.00895 Length: 2039
Score: 114.00 Matches: 31
Percent Similarity: 49.04% Conservative: 20
Best Local Similarity: 29.81% Mismatches: 47
Query Match: 5.57% Indels: 6
DB: 14 Gaps: 3

US-10-030-529A-1 (1-1168) x US-10-192-584-7 (1-2039)

QY 569 GAACGGTACTTATTAGATGATCTTATCGTATGATGAAACAAATACACATATATC 628
DB 1939 GlnLeuAsnThrValIleAspAsnValGlnAsnAsnPheAsnGlnValAsnGlnArgile 1958
QY 629 AATAAGTTGTCTAAGAAATGCAACTGTTTAGCCAAACCAATCAGCATTTGCTATGTTA 688
DB 1959 GlyAspLeuThrArgGluSerArgAlaGlyIleAlaGlyAlaMetAlaThrAlaSerLeu 1978
QY 689 GTCAACCAATGGTGTAGGCAAAACGAGCGTTTCTGCTCGGTAGGAGGTTATAGAGAT 748
DB 1979 GlnAsnValAlaLeuProGlyLysThrThrIleSerValGlyThrAlaThrPheLysGly 1998
QY 749 AAACGTGATTAGCATTTGTGCGCTACGCAATTACTGATCGTTTACCGCTAAAGCG 808
DB 1999 GlnAsnAlaValAlaIleGlyMet---SerArgLeuSerAsp-----AsnGlyLysVal 2015
QY 809 GGTGTA-----GGTTCAATACCTACAAATGCGGCATGCTTATGTTGCTGCTGTT 859
DB 2016 GlyIleArgLeuSerGlyMetSerThrSerAsnGlyAspLysGlyAlaAlaMetSerVal 2035
QY 860 GGTATGCAATTC 871
DB 2036 GlyPheThrPhe 2039

RESULT 8

US-09-952-267-13
; Sequence 13, Application US/09952267
; Publication No. US20030032772A1
; GENERAL INFORMATION:
; APPLICANT: HANSEN, ERIC J.
; APPLICANT: ABEI, CHRISTOPH
; APPLICANT: COPE, LESLIE D.
; APPLICANT: MACIVER, ISOBEL
; APPLICANT: FISKE, MICHAEL J.
; APPLICANT: FREDENBURG, ROSS A.

; TITLE OF INVENTION: USP41 AND USP42 ANTIGENS OF MORAXELLA CATARRHALIS
; FILE REFERENCE: AMCY:024
; CURRENT APPLICATION NUMBER: US/09/952,267
; CURRENT FILING DATE: 2001-09-12
; PRIOR APPLICATION NUMBER: 09/336,447
; PRIOR FILING DATE: 1999-06-21
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 13
; LENGTH: 873
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
; US-09-952-267-13

Alignment Scores:
Pred. No.: 0.019 Length: 873
Score: 110.00 Matches: 42
Percent Similarity: 47.37% Conservative: 30
Best Local Similarity: 27.63% Mismatches: 52
Query Match: 5.37% Indels: 28
DB: 10 Gaps: 8

US-10-030-529A-1 (1-1168) x US-09-952-267-13 (1-873)

QY 428 CTTGGAATAAATCGGCAGCAGCTGAAATTCATATAGTTATTTTAACGATTTAAGA 487
DB 745 IleAlaThrAsnLysGlnGluLeu-----IleLeuGln 755
QY 488 CACGATTTTAAATTAAGTCTCTGTATGACGCTATTTCCAAAATAAACAATATTGAT 547
DB 756 HisAsp---ArgLeuAsnArgileAsnGluThrAsnAsnArgGlnLysIleAsp 774
QY 548 ACTATAAGTAATATTTACTAGACCTGGGTACTTATTAGATGATCTTATCGTATGATG 607
DB 775 -----GlnLeuGly-----TyrAlaLeuLys 781
QY 608 GAACAAATACACATAATATCAATAGTTGTCTTAAA---GAATTCGCAACT-----GGT 658
DB 782 GluGlnGlyGlnHisPheAsnAsnArgIleSerAlaValGluArgGlnThrAlaGlyGly 801
QY 659 TTAGCCAAACCAATCAGCATTTGCTATGTTAGTGCAACCAATGGTGTAGGCAAAACGAGC 718
DB 802 IleAlaAsnAlaIleAlaIleAlaThrLeuProSerArgAlaGlyGluHisHis 821
QY 719 GTTCTGCTGCGGTAGGAGGTTATAGATATAAATCGCATTTAGCCATTTGGT---GTCCGC 775
DB 822 ValLeuPheGlySerGlyTyrHisAsnGlyGlnAlaAlaValSerLeuGlyAlaAlaGly 841
QY 776 TCACGCAATTACTGATCGCTTTACCGCTAAAGCGGGTGTAGCGTTCAATACCTACAATGGC 835
DB 842 LeuSerAspThrGlyLysSerThrTyrLysIleGlyLeuSerTrpSerAsp---AlaGly 860
QY 836 GGCATGCTCTTATGCTGCTTCTGTTGTTGTTATGATTC 871
DB 861 GlyLeuSerGlyGlyValGlyGlySerTyrArgTrp 872

RESULT 9

US-10-282-122A-62892
; Sequence 62892, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.

```
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 62892
; LENGTH: 852
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-10-282-122A-62892

Alignment Scores:
Pred. No.: 0.0308 Length: 852
Score: 108.00 Matches: 41
Percent Similarity: 48.73% Conservative: 36
Best Local Similarity: 25.95% Mismatches: 63
Query Match: 5.27% Indels: 18
DB: 12 Gaps: 8

US-10-030-529A-1 (1-1168) x US-10-282-122A-62892 (1-852)
QY 437 AATCGGCAGCAGTGAATTTGTTATAGTTATTTTAAACGATTTTACACAGCATTTT 496
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 699 AsnlystrAlaIleGluGlnAsnIleAsn-----ArgThrValAlaAsnGlyPhe 715
QY 497 AAATTAAGTCTTGATGTCAGCTATTTCACAAATAAATAACAAATATT----- 544
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 716 GluileGluLysAsnLysAlaGlyIleAlaThrAsnLysGlnGluLeuIleLeuGlnHis 735
QY 545 GATACTATAAGTAATATTACTAGAACTGGTACTTATTAGATGAT----- 592
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 736 AspArgLeuAsnArg---IleAsnGluThrAsnAsnHisGlnAspGlnLysIleAspGln 754
QY 593 ---TCTTATCGTATGATGGAACAAATACACATAATATCAATAGTTGTCATAA---GAA 646
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 755 LeuGlyTyrAlaLeuLysGluGlnGlyGlnHisPheAsnAsnArgIleSerAlaValGlu 774
QY 647 TTGCAAACT-----GGTTTAGCCCAACCAATACGATGTCTATGTTAGTGCACCAAAAT 700
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 775 ArgGlnThrAlaGlyIleAlaAsnAlaIleAlaThrLeuProSerProSer 794
QY 701 GGTGTAGCAAAACGCGGTTCTGCTCGCGTAGAGGTTATAGATATAAACTGCATT 760
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 795 ArgAlaGlyGluHisHisValLeuPheGlySerGlyTyrHisAsnGlyGlnAlaVal 814
QY 761 GCCATTGGT---GTCGGCTCACGCATTACTGATCGCTTACCGCTAAAGCGGTGTAGCG 817
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 815 SerLeuGlyAlaAlaGlyLeuSerAspThrGlyLysSerThrTyrLysIleGlyLeuSer 834
QY 818 TTCAATACCTCAATGCGCGCATGCTTATAGCTGCTTCTGTTGGTTATGAATTC 871
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

```
Db 835 TrpSerAsp---AlaGlyGlyLeuSerGlyValGlyGlySerTyrArgTrp 851
RESULT 10
US-09-952-267-5
; Sequence 5, Application US/09952267
; Publication No. US20030032772A1
; GENERAL INFORMATION:
; APPLICANT: HANSEN, ERIC J.
; APPLICANT: AEBI, CHRISTOPH
; APPLICANT: COPE, LESLIE D.
; APPLICANT: MACIVER, ISOBEL
; APPLICANT: FISKE, MICHAEL J.
; APPLICANT: FREDENBURG, ROSS A.
; TITLE OF INVENTION: USPA1 AND USPA2 ANTIGENS OF MORAXELLA CATARRHALIS
; FILE REFERENCE: AMCY.024
; CURRENT APPLICATION NUMBER: US/09/952,267
; CURRENT FILING DATE: 2001-09-12
; PRIOR APPLICATION NUMBER: 09/336,447
; PRIOR FILING DATE: 1999-06-21
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 892
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-09-952-267-5

Alignment Scores:
Pred. No.: 0.0311 Length: 892
Score: 108.00 Matches: 42
Percent Similarity: 46.71% Conservative: 29
Best Local Similarity: 27.63% Mismatches: 53
Query Match: 5.27% Indels: 28
DB: 10 Gaps: 8

US-10-030-529A-1 (1-1168) x US-09-952-267-5 (1-892)
QY 428 CTTGGAATAAATCGGCAGCAGCTGAAATTTGAATTTGTATAGTTATTTTAAACGATTTAAGA 487
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 764 IleAlaThrAsnLysGlnGluLeuIleGln-----AsnAsp----- 776
QY 488 CACGATTTTAAATTAAGTTCTTGATGTCAGCTATTTCACAAATAAATAACAAATATTGAT 547
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 777 -----ArgLeuAsnGlnIleAsnGluThrAsnAsnArgGlnAspGlnLysIleAsp 793
QY 548 ACTATAAGTAATATTACTAGAACTGGTACTTATTAGATGATCTTATCGTATGATG 607
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 794 -----GlnLeuGly-----TyrAlaLeuLys 800
QY 608 GAACAAATAACACATAATATCAATAAGTTGTCATAA---GAATTGCCAACT-----GGT 658
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 801 GluGlnGlyGlnHisPheAsnAsnArgIleSerAlaValGluArgGlnThrAlaGlyGly 820
QY 659 TTAGCAACCAACCAATGATGCTATGTTAGTGCACCAAAATGGTGTAGGCAAAACGAGC 718
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 821 IleAlaAsnAlaIleAlaThrLeuProSerProSerArgAlaGlyGluHisHis 840
QY 719 GTTTCGTGCTGGTAGGAGTTATAGATATAAACTGCATTAGCCATTGGT---GTCGCGC 775
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 841 ValLeuPheGlySerGlyTyrHisAsnGlyGlnAlaValSerLeuGlyAlaValGly 860
QY 776 TCACGCACTACTGATCGCTTTTACCGCTAAAGCGGTGTAGCGTTCAATACCTCAATAGC 835
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 861 LeuSerAspThrGlyLysSerThrTyrLysIleGlyLeuSerTrpSerAsp---AlaGly 879
QY 836 GCATGCTCTTAGGTGCTCTTCTGTTGGTTATGAATTC 871
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 880 GlyLeuSerGlyGlyValGlyGlySerTyrArgTrp 891
RESULT 11
US-09-952-267-1
; Sequence 1, Application US/09952267
; Publication No. US20030032772A1
```

```

; GENERAL INFORMATION:
; APPLICANT: HANSEN, ERIC J.
; APPLICANT: ABEI, CHRISTOPH
; APPLICANT: COPE, LESLIE D.
; APPLICANT: MACIVER, ISOBEL
; APPLICANT: FISKE, MICHAEL J.
; APPLICANT: FREDENBURG, ROSS A.
; TITLE OF INVENTION: USP1 AND USP2 ANTIGENS OF MORAXELLA CATARRHALIS
; CURRENT APPLICATION NUMBER: US/09/952,267
; PRIOR FILING DATE: 2001-09-12
; PRIOR FILING DATE: 1999-06-21
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 831
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-09-952-267-1

Alignment Scores:
Pred. No.: 0.039 Length: 831
Score: 107.00 Matches: 42
Percent Similarity: 46.71% Conservative: 29
Best Local Similarity: 27.63% Mismatches: 53
Query Match: 5.22% Indels: 28
DB: 10 Gaps: 8

US-10-030-529A-1 (1-1168) x US-09-952-267-1 (1-831)
QY 428 CTTGGAATAATCGGACGAGCTGAATTTGATAGTATTATTTTACGATTTAAGA 487
Db 703 ILEAlaThrAsnLysGlnGluLeuLeuLeuGln-----AsnAsp----- 715
QY 488 CACGATTTAAATTAAGTCTTGTATGACGATTTTCCAAAATAAACAATAATATGAT 547
Db 716 -----ArgLeuAsnArgIleAsnGluThrAsnAsnArgGlnAspGlnLysIleAsp 732
QY 548 ACTATAAGTAATATTACTAGAACCTGGGACTTTATTTAGATGATTTCTATCGTATG 607
Db 733 -----GlnLeuGly-----TyrAlaLeuLys 739
QY 608 GAACAAATACATAATATCAATAGTCTCTTAA-----CAATGCAAACT-----GGT 658
Db 740 GluGlnGlyGlnHisPheAsnAsnArgIleSerAlaValGluArgGlnThrAlaGlyGly 759
QY 659 TTAGCCAAACCAATCAGCATTTGCTATGTTAGTGCAACCAATGCTGTAGCAAAACGAGC 718
Db 760 ILEAlaAsnAlaIleAlaIleAlaThrLeuProSerProSerArgAlaGlyGluHisHis 779
QY 719 GTTCTGCTCGGTAGGAGGTTATAGAGATAAACTGCAATAGCCATTGGT---GTCGGC 775
Db 780 ValLeuPheGlySerGlyTyrHisAsnGlyGlnAlaValSerLeuGlyAlaAlaGly 799
QY 776 TCAGCATTAATGATCGCTTACCGTAAAGCGGTGTCGTTCACTCAATCAATGCG 835
Db 800 LeuSerAspThrGlyLysSerThrTyrLysIleGlyLeuSerTrpSerAsp---AlaGly 818
QY 836 GGCATGCTTATGCTGCTCTGTTGTTGTTATGAATTC 871
Db 819 GlyLeuSerGlyGlyValGlyGlySerTyrArgTrp 830

RESULT 12
US-09-952-267-9
; Sequence 9, Application US/09952267
; Publication No. US20030032772A1
; GENERAL INFORMATION:
; APPLICANT: HANSEN, ERIC J.
; APPLICANT: ABEI, CHRISTOPH
; APPLICANT: COPE, LESLIE D.
; APPLICANT: MACIVER, ISOBEL
; APPLICANT: FISKE, MICHAEL J.

; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIO
; APPLICANT: OCHIALI, KEIYO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES

; APPLICANT: FREDENBURG, ROSS A.
; TITLE OF INVENTION: USP1 AND USP2 ANTIGENS OF MORAXELLA CATARRHALIS
; FILE REFERENCE: AMCY:024
; CURRENT APPLICATION NUMBER: US/09/952,267
; CURRENT FILING DATE: 2001-09-12
; PRIOR FILING DATE: 1999-06-21
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 9
; LENGTH: 941
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-09-952-267-9

Alignment Scores:
Pred. No.: 0.0403 Length: 941
Score: 107.00 Matches: 41
Percent Similarity: 48.73% Conservative: 36
Best Local Similarity: 25.95% Mismatches: 63
Query Match: 5.22% Indels: 18
DB: 10 Gaps: 8

US-10-030-529A-1 (1-1168) x US-09-952-267-9 (1-941)
QY 437 AATCGGACGAGCTGAATTTGATAGTATTATTTTAAACGATTTAAGACAGCATTTT 496
Db 788 AsnLysThrAlaIleGluGlnAsnIleAsn-----ArgThrValAlaAsnGlyPhe 804
QY 497 AAATTAAGATTTCTGATGACGATTTTCCAAAATAAACAATAATATT----- 544
Db 805 GluIleGluLysAsnLysAlaGlyIleAlaThrAsnLysGlnGluLeuLeuGlnAsn 824
QY 545 GATACATATAAGTAATATTACTAGAACCTGGGACTTTATTTAGATGAT----- 592
Db 825 AspArgLeuAsnGln-----IleAsnGluThrAsnAsnHisGlnAspGlnLysIleAspGln 843
QY 593 ---TCTTATCGTATGAGCAACAAATAACATAATATCAATAGTCTCTTAA---GAA 646
Db 844 LeuGlyTyrAlaLeuLysGlnGlyGlnHisPheAsnAsnArgIleSerAlaValGlu 863
QY 647 TTCCAAACT-----GGTTAGCAACCAATCAGCATTTGCTATGTTAGTCAACCAAAAT 700
Db 864 ArgGlnThrAlaGlyGlyIleAlaAsnAlaIleAlaIleAlaThrLeuProSerProSer 883
QY 701 GGTGTAGGCAAAACGAGCGTTCTGTCGGGTAGGAGGTTATAGAGATAAACTGCATTA 760
Db 884 ArgAlaGlyGluHisValLeuPheGlySerGlyTyrHisAsnGlyGlnAlaAlaVal 903
QY 761 GCATTTGGT---GTCGGCTCAGCATTTACTGCTGCTTTTACCGCTAAAGCGGTGTAGCG 817
Db 904 SerLeuGlyAlaAlaGlyLeuSerAspThrGlyLysSerThrTyrLysIleGlyLeuSer 923
QY 818 TTCATATACCTACAATCGCGCATGCTCTTATGCTGCTCTTCTGTTGTTATGAATTC 871
Db 924 TrpSerAsp---AlaGlyGlyLeuSerGlyGlyValGlyGlySerTyrArgTrp 940

RESULT 13
US-09-738-626-5483
; Sequence 5483, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIO
; APPLICANT: OCHIALI, KEIYO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
```

```
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738.626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 5483
; LENGTH: 496
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-5483

Alignment Scores:
Pred. No.: 0.0627 Length: 496
Score: 104.50 Matches: 67
Percent Similarity: 40.00% Conservative: 43
Best Local Similarity: 24.36% Mismatches: 104
Query Match: 5.10% Indels: 61
DB: 9 Gaps: 16

US-10-030-529A-1 (1-1168) x US-09-738-626-5483 (1-496)
Qy 173 CCAAGTTTGGTGGAGTATCT-----TCTTGTATAGCTATGATGAC 217
Db 148 ProllysAlaAlaAspValSerGlnAlaThrValGlySerIleTySer----- 163
Qy 218 TATGTAAGGTAATAGGACTTGGCTAATGAAGC-----GGTTTCGATTTAAAGTG 271
Db 164 PheGlyAspGlyArgIleGlyLeuMetLeuGluAlaProGluGlyPheAspMetGlnVal 183
Qy 272 CCAGGGATTAATAAAGAAAG---CCAAAAAGATGGATTTCTAAACAGGCTACTTATCTTTGAA 328
Db 184 LeuAsnIleThrMetArgAlaLysLysGluLeuValGluValAspSerThrAsnLeuThr 203
Qy 329 TTACAGCATATATGCCCTTATCTACTCTCTCTCGTGCATATGCTCTCGGCTTCTCCT 388
Db 204 SerGluAsnMetLys-----AlaGlnGlyTyAspAsp 214
Qy 389 AGC-----CCTATCTGTTATATCCGATGCTCTGAT----- 418
Db 215 SerPheIleAspHisAspGlnTrpArgIleMetLeuAsnProAlaAlaGluThrLeuAla 234
Qy 419 -----CCTGATCAACTTGGAAATAATCGGCAGCAGCTGAAATTCGAATTTGTATAGTTAT 472
Db 235 ThrLeuPro-----AlaLeuAsnAspGluGlnIleAspLeuArgAlaLeu---Tyr 250
Qy 473 TTTAACGATTTAAGACACAGCATTTTAATTA-----AAAGTTCTTGATGCAGGTATT 523
Db 251 TrpGlnThrSerMetThrAspThrGluValGlyHisArgGlnAlaLeuAspThrSerLeu 270
Qy 524 TCCAAAAATAACAAAATAATGATGATATAAGTAAATATTTACTAGAACTGGGT----- 577
Db 271 SerArgMetAsnAlaLeuLeuGlyArgIleGluAsnTyLysGluAlaLeuGlyValThr 290
Qy 578 -----ACTTATTTAGATGATCTTATCGTATGATGAACAAAATAACATATATCAAT 631
Db 291 SerAlaThrIleMetGlyGluArgValSerLeuIleGluProIleProHis-----Glu 308
Qy 632 AAGTTGTCTAAGAAATGCAAACTGGTTTACCCAAACCAATCAGCATTTGCTATGTTAGTG 691
Db 309 SerIleArgGlyAspArgIleGluGlyLeuAlaSerValAspValThrLysGlnLeuAla 328
Qy 692 CAA---CCAAATGGGTAGGCAAAACGAGCGTTTCTGCTGCGGTAGGAGTTATAGAT 748
Db 329 GluAspProThrGly-----LysSerValProGlyTrpGluAsp 341
Qy 749 AAAACTGCATTAGCCATTGGTGTGCGCTCACGCAATCTACTGATCGCTTTTACCGCTAAAGCG 808
Db 749 AAAAGTTGTCTAAGAAATGCAAACTGGTTTACCCAAACCAATCAGCATTTGCTATGTTA 688

; 342 LysThrPheMetAlaAlaSerLeuAspAsnLeuMetSerGlu----- 355
Qy 809 GGTGTAGCGTTCAATACCTACCAATGCGCATGCTTATGCTGCTCTGTTGGTTATGAA 868
Db 356 GlyTyArgGlnTyAsnSerTyArgLeuValThrAspHisIlePheGlyGlyTyAsp 375
Qy 869 TTTCAATCAATACGTTTAATCACTAATCGTTTGGTTTATATAATAA 913
Db 376 Phe---AspTrpArgLeuGlyThrThrLeuAspGlyTyLeuLys 389

RESULT 14
US-09-797-862-32
; Sequence 32, Application US/09797862
; Patent No. US20020102276A1
; GENERAL INFORMATION:
; APPLICANT: PEAK, IAN RICHARD ANSELM
; APPLICANT: JENNINGS, MICHAEL PAUL
; APPLICANT: MOXON, E. RICHARD
; TITLE OF INVENTION: NOVEL SURFACE ANTIGEN
; FILE REFERENCE: 065064/0134
; CURRENT APPLICATION NUMBER: US/09/797,862
; CURRENT FILING DATE: 2001-05-03
; PRIOR APPLICATION NUMBER: PCT/AU98/01031
; PRIOR FILING DATE: 1998-12-14
; PRIOR APPLICATION NUMBER: GB 9726398.2
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 32
; LENGTH: 1098
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-09-797-862-32

Alignment Scores:
Pred. No.: 0.205 Length: 1098
Score: 100.50 Matches: 50
Percent Similarity: 32.88% Conservative: 23
Best Local Similarity: 22.52% Mismatches: 90
Query Match: 4.91% Indels: 59
DB: 9 Gaps: 5

US-10-030-529A-1 (1-1168) x US-09-797-862-32 (1-1098)
Qy 227 GGTAAATGGAGTCTGGTCTAATGAAGCGGTTTCGATATTAAGTCCAGGATTAATAATG 286
Db 929 GlyLysTrp-----TyrHisAlaLysAlaAspGlyThrAlaAsp 941
Qy 287 AAGCCAAAAGATGGATTTCTAAACAGGCTACTTATCTTGAATTCACGATTTATATGCCT 346
Db 942 LysThrLysGlyGluValSerAsnAspLysValSerThrAspGluLysHisValValSer 961
Qy 347 TATATCTCTGTTCTCGTGACATATGCTCCGCGGCTTCTCTAGCCCTATATCTGTATAT 406
Db 962 LeuAspProAsnAspGlnSerLysGlyValValIleAspAsnValAlaAsnGly 981
Qy 407 CCATGTCCTGATCTGATCAACTTGGAAATAATCGGCAGCAGCTG----- 451
Db 982 AspIleSerAlaThrSerThrAspAlaIleAsnGlySerGlnLeuTyAlaValAlaLys 1001
Qy 452 ---AAATTTGAATTTGTATAGTTATTTTAAACGATTTAAACACACGATTTTAAATTTAAAGTT 508
Db 1002 GlyValThrAsnLeuAlaGlyGlnValAsnAsnLeuGlu----- 1014
Qy 509 CTTGATCGCATGATTTCCAAAAATAACAAATAATTGATCTATATAAGTAATATTTACTA 568
Db 1014 ----- 1014
Qy 569 GAACCTGGTACTTATTATTAGATGATTTCTTATCGTATGATGGAACAAATACACATAATATC 628
Db 1015 -----GlyLysVal 1017
Qy 629 AATAAGTTGTCTAAGAAATGCAAACTGGTTTACCCAAACCAATCAGCATTTGCTATGTTA 688
```

Db	:::			:::						1037
1018	AsnLysValGlyLysArgAlaAspAlaGlyThrAlaSerAlaLeuAlaAlaSerGlnLeu									1037
Qy	GTCCAAACCAATGGTGTAGGCAAAACAGAGCGTTTCTCTCGCGTAGGAGGTTTATAGAGAT									748
Db	:::			:::						1057
1038	ProGlnAlaThrMetProGlyLysSerMetValAlaIleAlaGlySerSerTyrGlnGly									1057
Qy	AAAAATGCATTAGGCATTGGTGTGCGCTCACGCATTACTCATCGCTTACCGCTTAAGCG									808
Db	:::			:::						1076
1058	GlnAsnGlyLeuAlaIleGlyVal---SerArgIleSerAspAsnGlyLysValIleIle									1076
Qy	GGGTGACGGTTCATACCTACATGGCGCATGTCT---TATGGTGCTTCTGTGGTTAT									865
Db	:::			:::						1096
1077	ArgLysSerGlyThrThrAsnSerGlnGlyLysThrGlyValAlaAlaGlyValGlyTyr									1096
Qy	GAATTC	871								
Db										1098
1097	GlnTirp	1098								

RESULT 15

US-09-797-862-33
; Sequence 33, Application US/09797862
; Patent No. US20020102276A1
; GENERAL INFORMATION:
; APPLICANT: PEAK, IAN RICHARD ANSELM
; APPLICANT: JENNINGS, MICHAEL PAUL
; APPLICANT: MOXON, E. RICHARD

US-10-030-529A-1 (1-1168) x US-09-797-862-33 (1-2353)

Qy	221	GGTAAGGGTAAATGCAGCTGGTCTCAAT--GAAGGGGGTTTCGATATTAAGTGCACGG	2177
Db	2182	GlyAlaAspLysTrpTyrtYrThrAsnAlaAspGlyAlaAlaAsp	2196
Qy	278	ATTAATAATGAAGCCAAAGAATGGATTCTTAAACAGGCTACTTATCTTGAATTACAGCAT	337
Db	2197	-----LysThrLysGlyGluValSerAsnAspLysValSerThrAspGluLysHis	2213
Qy	338	TATATGCTTTATACTCTGTCTCTGTCGACATATGCTCTCGCGCTTCTCTAGTCCCTATA	397
		:::	
Db	2214	ValValArgLeuAspProAsnAsnGlnSerAsnGlyLysGlyValValIleAspAsnVal	2233
Qy	398	CTGTATATCCGATGCTCGATCTCTGATCAACTTGGATAAATCGGCAGCAGCTG----	451
Db	2234	AlaAsnGlyGluIleSerAlaIaThrSerThrAspAlaIleAsnGlySerGlnLeuTyrAla	2253
Qy	452	-----AAATTGAATTTGTATAGTTATTTTAAACGATTTAAGACACGAGTTTAA	499
Db	2254	ValAlaLysGlyValThrAsnLeuAlaGlyGlnValAsnAsnLeuGlu-----	2269

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: May 13, 2004, 08:22:52 ; Search time 20 Seconds
(without alignments)
6029.913 Million cell updates/sec

Title: US-10-030-529A-1
Perfect score: 2048
Sequence: 1 ataataacgtcaatgacatt.....aagccgttaaaagtgcggac 1168

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 389414 seqs, 51625971 residues
Total number of hits satisfying chosen parameters: 778828

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO.spool.P/US10030529/runat 13052004 081453 2656/app_query.fasta_1.1351
-DB=Issued Patents AA -QFMT=fastan -SUFFIX=rai -MINMATCH=0.1 -LOOPL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10030529 @CGN 1 1 27 @runat 13052004 081453 2656 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents AA.*
1: /cgn2_6/ptodata/2/iaa/5A COMB.pcp.*
2: /cgn2_6/ptodata/2/iaa/5B COMB.pcp.*
3: /cgn2_6/ptodata/2/iaa/6A COMB.pcp.*
4: /cgn2_6/ptodata/2/iaa/6B COMB.pcp.*
5: /cgn2_6/ptodata/2/iaa/PCTUS COMB.pcp.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pcp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	193	9.4	610	4	US-09-336-447A-11
2	193	9.4	624	4	US-09-336-447A-7
3	192	9.4	889	4	US-09-336-447A-15
4	192	9.4	867	4	US-09-540-236-2676
5	190.5	9.3	573	4	US-09-336-447A-3
6	134	6.7	436	4	US-09-489-037A-11785
7	123.5	6.1	568	4	US-09-543-681A-6966
8	115	5.6	2042	4	US-09-077-098A-6
9	114	5.6	2039	4	US-09-077-098A-7
10	110	5.4	873	4	US-09-336-447A-13
11	108	5.3	892	4	US-09-336-447A-5
12	107	5.2	831	4	US-09-336-447A-1

13	107	5.2	878	4	US-09-540-236-3401	Sequence 3401, Ap
14	107	5.2	941	4	US-09-336-447A-9	Sequence 9, Appl
15	104	5.1	1002	4	US-09-268-347-24	Sequence 24, Appl
16	104	5.1	1004	4	US-09-268-347-30	Sequence 30, Appl
17	103.5	5.1	616	4	US-09-268-347-38	Sequence 38, Appl
18	103	5.0	1104	4	US-09-268-347-28	Sequence 28, Appl
19	103	5.0	1104	4	US-09-268-347-34	Sequence 34, Appl
20	102	5.0	679	3	US-08-913-942-15	Sequence 15, Appl
21	102	5.0	679	3	US-09-268-347-26	Sequence 26, Appl
22	102	5.0	1094	4	US-09-268-347-32	Sequence 32, Appl
23	101	4.9	1690	4	US-09-595-684B-39	Sequence 39, Appl
24	100.5	4.9	1098	1	US-08-409-995-2	Sequence 2, Appl
25	100.5	4.9	1098	3	US-08-685-467-2	Sequence 2, Appl
26	100.5	4.9	1098	3	US-09-377-155-32	Sequence 32, Appl
27	100.5	4.9	1098	3	US-08-913-942-2	Sequence 2, Appl
28	100.5	4.9	1098	4	US-09-669-974-32	Sequence 32, Appl
29	100.5	4.9	1098	4	US-09-268-347-44	Sequence 44, Appl
30	100.5	4.9	1098	4	US-09-797-862-32	Sequence 32, Appl
31	100.5	4.9	2353	3	US-09-377-155-33	Sequence 33, Appl
32	100.5	4.9	2353	3	US-08-913-942-4	Sequence 4, Appl
33	100.5	4.9	2353	4	US-09-669-974-33	Sequence 33, Appl
34	100.5	4.9	2353	4	US-09-797-862-33	Sequence 33, Appl
35	100.5	4.9	2411	4	US-09-268-347-36	Sequence 36, Appl
36	99	4.8	2354	4	US-09-268-347-47	Sequence 47, Appl
37	93	4.7	431	4	US-09-543-681A-5397	Sequence 5397, Ap
38	93	4.5	1117	2	US-08-843-530B-33	Sequence 33, Appl
39	92	4.5	447	4	US-09-543-681A-6231	Sequence 6231, Ap
40	90.5	4.4	1226	1	US-08-280-443-2	Sequence 2, Appl
41	90.5	4.4	1226	1	US-08-457-459-2	Sequence 2, Appl
42	90.5	4.4	1226	1	US-08-555-678-2	Sequence 2, Appl
43	90.5	4.4	1226	5	PCT-US95-02275-2	Sequence 2, Appl
44	90	4.4	512	4	US-09-489-039A-12836	Sequence 12836, A
45	90	4.4	944	4	US-09-134-001C-4352	Sequence 4352, Ap

ALIGNMENTS

RESULT 1
US-09-336-447A-11
; Sequence 11, Application US/09336447A
; Patent No. 6310190
; GENERAL INFORMATION:
; APPLICANT: HANSEN, ERIC J.
; APPLICANT: ABEI, CHRISTOPH
; APPLICANT: COPE, LESLIE D.
; APPLICANT: MACIVER, ISOBEL
; APPLICANT: FISKE, MICHAEL J.
; APPLICANT: FREDENBURG, ROSS A.
; TITLE OF INVENTION: USPA1 AND USPA2 ANTIGENS OF MORAXELLA CATARRHALIS
; FILE REFERENCE: AWCY:024
; CURRENT APPLICATION NUMBER: US/09/336,447A
; CURRENT FILING DATE: 1999-06-21
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 11
; LENGTH: 610
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-09-336-447A-11

Alignment Scores:
Pred. No.: 3.6e-13 Length: 610
Score: 193.00 Matches: 53
Percent Similarity: 50.71% Conservative: 18
Best Local Similarity: 37.86% Mismatches: 61
Query Match: 9.42% Indels: 8
DB: 4 Gaps: 3

US-10-030-529A-1 (1-1168) x US-09-336-447A-11 (1-610)

Qy 476 AACGATTTAAGACACGATTTTAAATTAAGTTCTTGATCGACGATTTTCAAAATAAAA 535

Db 471 AsnLysAlaSerAlaAspThrLysPheAlaAlaThrAlaAspAlaIleThrLysAsnGly 490

```
Qy 536 CAAATATTGATCTACTATAAGTAAATATTACTAGAACTGGGTACTTATTAGATGAT--- 592
Db 491 AsnAlaIleThrLysAsnAlaLysSerIleThrAspLeuGlyThrLysValAspGlyPhe 510
Qy 593 TCTTATCGTATGATGGAACAAATACACAT-----AATATCAATAAG 634
Db 511 AspGlyArgValThrAlaLeuAspThrLysValAsnAlaPheAspGlyArgIleThrAla 530
Qy 635 TTGCTAAAGAAATGCAAACTGGTTAGCCAAACCAATCAGCATGCTCTATGTTAGTGCAA 694
Db 531 LeuAspSerLysValGluAsnGlyMetAlaAlaGlnAlaLeuSerGlyLeuPheGln 550
Qy 695 CCAATGCTAGGCAAAACGAGGCTTCTGCTCGGTAGGAGGTATTAGAGATAAAACT 754
Db 551 ProTyrSerValGlyLysPheAsnAlaThrAlaAlaLeuGlyThrGlySerLysSer 570
Qy 755 GCATTAGCCATTGGTGTGCGTCAAGCATTTACTGATCGCTTTACCGCTAAAGCGGGTGA 814
Db 571 AlaValAlaIleGlyAlaGlyTyrArgValAsnProAsnLeuAlaPheLysAlaGlyAla 590
Qy 815 GCGTTCATACCTAC---AATGGCGGCATGCTTATGGTGTCTCTGTTGGTTATGAATTC 871
Db 591 AlaIleAsnThrSerGlyAsnLysLysGlySerTyrAsnIleGlyValAsnTyrGluPhe 610

RESULT 2
US-09-336-447A-7
; Sequence 7, Application US/09336447A
; Patent No. 6310190
; GENERAL INFORMATION:
; APPLICANT: HANSEN, ERIC J.
; APPLICANT: ABEI, CHRISTOPH
; APPLICANT: COPE, LESLIE D.
; APPLICANT: MACIVER, ISOBEL
; APPLICANT: FISKE, MICHAEL J.
; TITLE OF INVENTION: USP1 AND USP2 ANTIGENS OF MORAXELLA CATARRHALIS
; FILE REFERENCE: AMCY:024
; CURRENT APPLICATION NUMBER: US/09/336,447A
; CURRENT FILING DATE: 1999-06-21
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 7
; LENGTH: 624
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-09-336-447A-7

Alignment Scores:
Pred. No.: 3.63e-13 Length: 624
Score: 193.00 Matches: 53
Percent Similarity: 50.71% Conservative: 18
Best Local Similarity: 37.86% Mismatches: 61
Query Match: 9.42% Indels: 8
DB: 4 Gaps: 3

US-10-030-529A-1 (1-1168) x US-09-336-447A-7 (1-624)
Qy 476 AACGATTTAAGACACGAGTATTTAAATTAAGATCTTGATGACGATTTCCAAAATAAA 535
Db 485 AsnLysAlaSerAlaAspThrLysPheAlaAlaThrAlaAspAlaIleThrLysAsnGly 504
Qy 536 CAAATATTGATCTACTATAAGTAAATATTACTAGAACTGGGTACTTATTAGATGAT--- 592
Db 505 AsnAlaIleThrLysAsnAlaLysSerIleThrAspLeuGlyThrLysValAspGlyPhe 524
Qy 593 TCTTATCGTATGGAACAAATACACAT-----AATATCAATAAG 634
Db 525 AspGlyArgValThrAlaLeuAspThrLysValAsnAlaPheAspGlyArgIleThrAla 544
Qy 635 TTGCTAAAGAAATGCAAACTGGTTAGCCAAACCAATCAGCATGCTCTATGTTAGTGCAA 694
Db 545 LeuAspSerLysValGluAsnGlyMetAlaAlaGlnAlaLeuSerGlyLeuPheGln 564
```

```
Qy 695 CCAATGCTAGGCAAAACGAGCGTTTCTGCTCGGTAGGAGGTATTAGAGATAAAACT 754
Db 565 ProTyrSerValGlyLysPheAsnAlaThrAlaAlaLeuGlyThrGlySerLysSer 584
Qy 755 GCATTAGCCATTGGTGTGCGCTCAGCATTTACTGATCGCTTTACCGCTAAAGCGGGTGA 814
Db 585 AlaValAlaIleGlyAlaGlyTyrArgValAsnProAsnLeuAlaPheLysAlaGlyAla 604
Qy 815 GCGTTCATACCTAC---AATGGCGGCATGCTTATGGTGTCTCTGTTGGTTATGAATTC 871
Db 605 AlaIleAsnThrSerGlyAsnLysLysGlySerTyrAsnIleGlyValAsnTyrGluPhe 624

RESULT 3
US-09-336-447A-15
; Sequence 15, Application US/09336447A
; Patent No. 6310190
; GENERAL INFORMATION:
; APPLICANT: HANSEN, ERIC J.
; APPLICANT: ABEI, CHRISTOPH
; APPLICANT: COPE, LESLIE D.
; APPLICANT: MACIVER, ISOBEL
; APPLICANT: FISKE, MICHAEL J.
; TITLE OF INVENTION: USP1 AND USP2 ANTIGENS OF MORAXELLA CATARRHALIS
; FILE REFERENCE: AMCY:024
; CURRENT APPLICATION NUMBER: US/09/336,447A
; CURRENT FILING DATE: 1999-06-21
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 15
; LENGTH: 889
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-09-336-447A-15

Alignment Scores:
Pred. No.: 4.15e-13 Length: 889
Score: 193.00 Matches: 53
Percent Similarity: 50.71% Conservative: 18
Best Local Similarity: 37.86% Mismatches: 61
Query Match: 9.42% Indels: 8
DB: 4 Gaps: 3

US-10-030-529A-1 (1-1168) x US-09-336-447A-15 (1-889)
Qy 476 AACGATTTAAGACACGAGTATTTAAATTAAGATCTTGATGACGATTTCCAAAATAAA 535
Db 750 AsnLysAlaSerAlaAspThrLysPheAlaAlaThrAlaAspAlaIleThrLysAsnGly 769
Qy 536 CAAATATTGATCTACTATAAGTAAATATTACTAGAACTGGGTACTTATTAGATGAT--- 592
Db 770 AsnAlaIleThrLysAsnAlaLysSerIleThrAspLeuGlyThrLysValAspGlyPhe 789
Qy 593 TCTTATCGTATGGAACAAATACACAT-----AATATCAATAAG 634
Db 790 AspGlyArgValThrAlaLeuAspThrLysValAsnAlaPheAspGlyArgIleThrAla 809
Qy 635 TTGCTAAAGAAATGCAAACTGGTTAGCCAAACCAATCAGCATGCTCTATGTTAGTGCAA 694
Db 810 LeuAspSerLysValGluAsnGlyMetAlaAlaGlnAlaLeuSerGlyLeuPheGln 829
Qy 695 CCAATGCTAGGCAAAACGAGCGTTTCTGCTCGGTAGGAGGTATTAGAGATAAAACT 754
Db 830 ProTyrSerValGlyLysPheAsnAlaThrAlaAlaLeuGlyThrGlySerLysSer 849
Qy 755 GCATTAGCCATTGGTGTGCGCTCAGCATTTACTGATCGCTTTACCGCTAAAGCGGGTGA 814
Db 850 AlaValAlaIleGlyAlaGlyTyrArgValAsnProAsnLeuAlaPheLysAlaGlyAla 869
Qy 815 GCGTTCATACCTAC---AATGGCGGCATGCTTATGGTGTCTCTGTTGGTTATGAATTC 871
Db 870 AlaIleAsnThrSerGlyAsnLysLysGlySerTyrAsnIleGlyValAsnTyrGluPhe 889
```



```

; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 573
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-09-336-447A-3

Alignment Scores:
Pred. No.: 6.76e-13 Length: 573
Score: 190.50 Matches: 52
Percent Similarity: 47.62% Conservative: 18
Best Local Similarity: 35.37% Mismatches: 62
Query Match: 9.30% Indels: 15
DB: 4 Gaps: 3

US-10-030-529A-1 (1-1168) x US-09-336-447A-3 (1-573)
QY 476 AACGGATTAAAGACACAGCATTTTAAATTTAAAGTTCTTGATGCACGTTATTTCACAAAATAAA 535
Db AsnLysAlaSerAlaAspThrLysPheAlaAlaThrAlaAspAlaIleThrLysAsnGly 446
QY 536 CAAATAATTCATCTACTATAAGTAAATATTTACTAGAACTGGGT----- 577
Db AsnAlaIleThrLysAsnAlaLysSerIleThrAspLeuGlyThrLysValAspGlyPhe 466
QY 578 -----ACTTATTTAGATGATCTTATCGTATGATGGAACAAAAATACACAT--- 622
Db AspGlyArgValThrAlaLeuAspThrLysValAsnAlaLeuAspThrLysValAsnAla 486
QY 623 -----AATATCAATAGTTCTCTAAAGAANTGCAAACTGGTTAGCCCAACCAATCA 673
Db PheAspGlyArgIleThrAlaLeuAspSerLysValGluAsnGlyMetAlaAlaGlnAla 506
QY 674 GCATTGTCTATGTTAGTGCACCAAAATGGTGTAGGCAAAACGAGCGTTTCTGCTGCGGTA 733
Db AlaLeuSerGlyLeuPheGlnProTyrSerValGlyLysPheAsnAlaThrAlaAlaLeu 526
QY 734 GGAGGTTATAGAGATAAAACTGCANTTAGCCATTGGTGTGCGCTCACGCATTACTGATCGC 793
Db GlyGlyTyrGlySerLysSerAlaValAlaIleGlyAlaGlyTyrArgValAsnProAsn 546
QY 794 TTATACCGCTAAAGCGGTGTAGCGTTCAATACCTAC---AATGCGGCGCATGCTTATGCT 850
Db LeuAlaPheLysAlaGlyAlaAlaIleAsnThrSerGlyAsnLysLysGlySerTyrAsn 566
QY 851 GCTTCTGTTGGTTATGAATTC 871
Db IleGlyValAsnTyrGluPhe 573

RESULT 6
US-09-489-039A-11785
; Sequence 11785, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709, 2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 1432
; SEQ ID NO 11785
; LENGTH: 436
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-11785

Alignment Scores:
Pred. No.: 1.58e-06 Length: 436
Score: 134.00 Matches: 25
Percent Similarity: 81.40% Conservative: 10

```

[illegible]

Db 1942 GlnLeuAsnThrValIleAspAsnValGlnAsnAsnPheAsnGlnValAsnGlnArgIle 1961
Qy 629 AATAAGTTCTAAGAATTCGAACCTGGTTAGCCCAACCAATCAGCATTGCTATGTTA 688
Db 1962 GlyAspLeuThrArgGluSerArgAlaGlyIleAlaGlyAlaMetAlaThrAlaSerLeu 1981
Qy 689 GTGCAACCAATGCTAGGCAAAACGAGCGTTTCTGCTCGGTAGGAGGTATAGAGAT 748
Db 1982 GlnAsnValAlaLeuProGlyLysThrIleSerValGlyThrAlaThrPheLysGly 2001
Qy 749 AAACTGATAGCATTGGTGTGGCTCAGCATTACTGATCGCTTTACCGCTAAAGCG 808
Db 2002 GluAsnAlaValAlaIleGlyMet---SerArgLeuSerAsp-----AsnGlyLysVal 2018
Qy 809 GGTGTA-----CGCTCAATACCTACAAATCGCGGCATGCTTATGCTGCTCTGTT 859
Db 2019 GlyIleArgLeuSerGlyMetSerThrSerAsnGlyAspLysGlyAlaAlaMetSerVal 2038
Qy 860 GGTATGAATTC 871
Db 2039 GlyPheSerPhe 2042

RESULT 9

US-09-077-098A-7
; Sequence 7, Application US/09077098A
; Patent No. 6544519

GENERAL INFORMATION:

APPLICANT: TOKUNAGA, Eiji
SAKAGUCHI, Masashi
MATSUO, Kazuo
HAMADA, Fukusaburo

TOKIYOSHI, Sachio

PARAGALLINARUM AND PROCESS FOR PREPARING THE SAME

TITLE OF INVENTION: NOVEL POLYPEPTIDE FROM HAEMOPHILUS

NUMBER OF SEQUENCES: 8

CORRESPONDENCE ADDRESS:

ADDRESSEE: BROWDY AND NEIMARK

STREET: 624 Ninth Street, N.W., Suite 300

CITY: Washington

STATE: D.C.

COUNTRY: USA

ZIP: 20001

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/077,098A

FILING DATE: 19-May-1998

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/JP97/03222

FILING DATE: 12-SEP-1997

APPLICATION NUMBER: JP 27,148/1996

FILING DATE: 19-SEP-1996

ATTORNEY/AGENT INFORMATION:

NAME: KORNEAU, Anne M.

REGISTRATION NUMBER: 25,618

REFERENCE/DOCKET NUMBER: TOKUNAGA-1

TELECOMMUNICATION INFORMATION:

TELEPHONE: 202-628-5197

TELEFAX: 202-737-3528

INFORMATION FOR SEQ ID NO: 7:

SEQUENCE CHARACTERISTICS:

LENGTH: 2039 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

SEQUENCE DESCRIPTION: SEQ ID NO: 7:

US-09-077-098A-7

Alignment Scores:

Pred. No.: 0.000525 Length: 2039
Score: 114.00 Matches: 31
Percent Similarity: 49.04% Conservative: 20
Best Local Similarity: 29.81% Mismatches: 47
Query Match: 5.57% Indels: 6
DB: 4 Gaps: 3

US-10-030-529A-1 (1-1168) x US-09-077-098A-7 (1-2039)

Qy 569 GAACCTGGTACTTATTAGATGATCTTATCGTATGATGAAACAAATACACATAATATC 628
Db 1939 GlnLeuAsnThrValIleAspAsnValGlnAsnAsnPheAsnGlnValAsnGlnArgIle 1958
Qy 629 AATAAGTTCTAAGAATTCGAACCTGGTTAGCCCAACCAATCAGCATTGCTATGTTA 688
Db 1959 GlyAspLeuThrArgGluSerArgAlaGlyIleAlaGlyAlaMetAlaThrAlaSerLeu 1978
Qy 689 GTGCAACCAATGCTAGGCAAAACGAGCGTTTCTGCTCGGTAGGAGGTATAGAGAT 748
Db 1979 GlnAsnValAlaLeuProGlyLysThrIleSerValGlyThrAlaThrPheLysGly 1998
Qy 749 AAACTGATAGCATTGGTGTGGCTCAGCATTACTGATCGCTTTACCGCTAAAGCG 808
Db 1999 GluAsnAlaValAlaIleGlyMet---SerArgLeuSerAsp-----AsnGlyLysVal 2015
Qy 809 GGTGTA-----CGCTCAATACCTACAAATCGCGGCATGCTTATGCTGCTCTGTT 859
Db 2016 GlyIleArgLeuSerGlyMetSerThrSerAsnGlyAspLysGlyAlaAlaMetSerVal 2035
Qy 860 GGTATGAATTC 871
Db 2036 GlyPheThrPhe 2039

RESULT 10

US-09-336-447A-13
; Sequence 13, Application US/09336447A
; Patent No. 6310190

GENERAL INFORMATION:

APPLICANT: HANSEN, ERIC J.

APPLICANT: AEBI, CHRISTOPH

APPLICANT: COPE, LESLIE D.

APPLICANT: MACIVER, ISOBEL

APPLICANT: FISKE, MICHAEL J.

APPLICANT: FREDENBURG, ROSS A.

TITLE OF INVENTION: USP2 ANTIGENS OF MORAXELLA CATARRHALIS

FILE REFERENCE: AMCY:024

CURRENT FILING DATE: 1999-06-21

NUMBER OF SEQ ID NOS: 98

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 13

LENGTH: 873

TYPE: PRT

ORGANISM: Moraxella catarrhalis

US-09-336-447A-13

Alignment Scores:

Pred. No.: 0.00109 Length: 873
Score: 110.00 Matches: 42
Percent Similarity: 47.37% Conservative: 30
Best Local Similarity: 27.63% Mismatches: 52
Query Match: 5.37% Indels: 28
DB: 4 Gaps: 8

US-10-030-529A-1 (1-1168) x US-09-336-447A-13 (1-873)

Qy 428 CTTGGATTAATCGGCGAGCGAGCTGAATTCGAATTTGTATAGTTATTTAACGATTTAAGA 487
Db 745 IleAlaThrAsnLysGlnGluLeu-----IleLeuGln 755
Qy 488 CAGGATTTAAATTAAGGTTCTTGATGACGATTTTCCAAATAAACAATAATTGAT 547
Db 756 HisAsp---ArgLeuAsnArgIleAsnGluThrAsnAsnArgGlnAspGlnLysIleAsp 774

QY 548 ACTATAAGTAAATATTTACTAGAACTGGGTACTTATTATAGATGATTTCTTATCGTATGATG 607
Db :|||:|||||
775 -----GlnLeuGly-----TyrAlaLeuLys 781
QY 608 GAACAAAATACACATAATATCAATAAGTTGTCTAAA---GAATTGCAAACT-----GGT 658
Db |||:|||||
782 GluGlnGlyGlnHisPheAsnAsnArgIleSerAlaValGluArgGlnThrAlaGlyGly 801
QY 659 TTAGCCAAACCAATCAGCATTTCTCTATGTTAGTGCACCAAACTGGGTAGGCAAAACGAGC 718
Db :|||:|||||
802 IleAlaAsnAlaIleAlaIleAlaThrLeuProSerProSerArgAlaGlyGluHis 821
QY 719 GTTCTCTGCTGCGGTAGGTTATAGAGATAAACTGCATTAGCCATTGGT---GTGCGC 775
Db |||:|||||
822 ValLeuPheGlySerGlyTyrHisAsnGlyGlnAlaValSerLeuGlyAlaAlaGly 841
QY 776 TCACCATTTACTGATCGCTTTTACCCTAAAGCGGTGTAGCGTTCAATCACTCAATGCGC 835
Db |||:|||||
842 LeuSerAspThrGlyLysSerThrTyrLysIleGlyLeuSerTrpSerAsp---AlaGly 860
QY 836 GGCATGTTCTATGGTGCTTCTGTTGGTTATGNAATTC 871
Db |||:|||||
861 GlyLeuSerGlyGlyValGlyGlySerTyrArgTrp 872

RESULT 11

US-09-336-447A-5
; Sequence 5, Application US/09336447A
; Patent No. 6310190
; GENERAL INFORMATION:
; APPLICANT: HANSEN, ERIC J.
; APPLICANT: AEBI, CHRISTOPH
; APPLICANT: COPE, LESLIE D.
; APPLICANT: MACIVER, ISOBEL
; APPLICANT: FISKE, MICHAEL J.
; APPLICANT: FREDENBURG, ROSS A.
; TITLE OF INVENTION: USP1 AND USP2 ANTIGENS OF MORAXELLA CATARRHALIS
; FILE REFERENCE: AMCY:024
; CURRENT APPLICATION NUMBER: US/09/336,447A
; CURRENT FILING DATE: 1999-06-21
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-09-336-447A-5

Alignment Scores:
Pred. No.: 0.00185 Length: 892
Score: 108.00 Matches: 42
Percent Similarity: 46.71% Conservative: 29
Best Local Similarity: 27.63% Mismatches: 53
Query Match: 5.27% Indels: 28
DB: 4 Gaps: 8

US-10-030-529A-1 (1-1168) x US-09-336-447A-5 (1-892)

QY 428 CTTGGAATAAATCGGCAGCAGCTGAAATTTGATAGTATTATTAACGATTAAAGA 487
Db :|||:|||||
764 IleAlaThrAsnLysGlnLeuIleLeuGln-----AsnAsp----- 776
QY 488 CACGATTTTAAATTAAGTTCTTGTGTCACGCTATTTCACAAAATAACAAATATTGAT 547
Db :|||:|||||
777 -----ArgLeuAsnGlnIleAsnGluThrAsnAsnArgGlnAspGlnLysIleAsp 793
QY 548 ACTATAAGTAAATATTTACTAGAACTGGGTACTTATTATAGATGATTTCTTATCGTATGATG 607
Db :|||:|||||
794 -----GlnLeuGly-----TyrAlaLeuLys 800
QY 608 GAACAAAATACACATAATATCAATAAGTTGTCTAAA---GAATTGCAAACT-----GGT 658
Db |||:|||||
801 GluGlnGlyGlnHisPheAsnAsnArgIleSerAlaValGluArgGlnThrAlaGlyGly 820

QY 659 TTAGCCAAACCAATCAGCATTTCTCTATGTTAGTGCACCAAACTGGGTAGGCAAAACGAGC 718
Db :|||:|||||
821 IleAlaAsnAlaIleAlaIleAlaThrLeuProSerProSerArgAlaGlyGluHis 840
QY 719 GTTCTCTGCTGCGGTAGGTTATAGAGATAAACTGCATTAGCCATTGGT---GTGCGC 775
Db |||:|||||
841 ValLeuPheGlySerGlyTyrHisAsnGlyGlnAlaValSerLeuGlyAlaAlaGly 860
QY 776 TCACCATTTACTGATCGCTTTTACCCTAAAGCGGTGTAGCGTTCAATCACTCAATGCGC 835
Db :|||:|||||
861 LeuSerAspThrGlyLysSerThrTyrLysIleGlyLeuSerTrpSerAsp---AlaGly 879
QY 836 GGCATGTTCTATGGTGCTTCTGTTGGTTATGNAATTC 871
Db |||:|||||
880 GlyLeuSerGlyGlyValGlyGlySerTyrArgTrp 891

RESULT 12

US-09-336-447A-1
; Sequence 1, Application US/09336447A
; Patent No. 6310190
; GENERAL INFORMATION:
; APPLICANT: HANSEN, ERIC J.
; APPLICANT: AEBI, CHRISTOPH
; APPLICANT: COPE, LESLIE D.
; APPLICANT: MACIVER, ISOBEL
; APPLICANT: FISKE, MICHAEL J.
; APPLICANT: FREDENBURG, ROSS A.
; TITLE OF INVENTION: USP1 AND USP2 ANTIGENS OF MORAXELLA CATARRHALIS
; FILE REFERENCE: AMCY:024
; CURRENT APPLICATION NUMBER: US/09/336,447A
; CURRENT FILING DATE: 1999-06-21
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 831
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-09-336-447A-1

Alignment Scores:
Pred. No.: 0.00234 Length: 831
Score: 107.00 Matches: 42
Percent Similarity: 46.71% Conservative: 29
Best Local Similarity: 27.63% Mismatches: 53
Query Match: 5.22% Indels: 28
DB: 4 Gaps: 8

US-10-030-529A-1 (1-1168) x US-09-336-447A-1 (1-831)

QY 428 CTTGGAATAAATCGGCAGCAGCTGAAATTTGATAGTATTATTAACGATTAAAGA 487
Db :|||:|||||
703 IleAlaThrAsnLysGlnGluLeuIleLeuGln-----AsnAsp----- 715
QY 488 CACGATTTTAAATTAAGTTCTTGTGTCACGCTATTTCACAAAATAACAAATATTGAT 547
Db :|||:|||||
716 -----ArgLeuAsnArgIleAsnGluThrAsnAsnArgGlnAspGlnLysIleAsp 732
QY 548 ACTATAAGTAAATATTTACTAGAACTGGGTACTTATTATAGATGATTTCTTATCGTATGATG 607
Db :|||:|||||
733 -----GlnLeuGly-----TyrAlaLeuLys 739
QY 608 GAACAAAATACACATAATATCAATAAGTTGTCTAAA---GAATTGCAAACT-----GGT 658
Db |||:|||||
740 GluGlnGlyGlnHisPheAsnAsnArgIleSerAlaValGluArgGlnThrAlaGlyGly 759
QY 659 TTAGCCAAACCAATCAGCATTTCTCTATGTTAGTGCACCAAACTGGGTAGGCAAAACGAGC 718
Db :|||:|||||
760 IleAlaAsnAlaIleAlaIleAlaThrLeuProSerProSerArgAlaGlyGluHis 779
QY 719 GTTCTCTGCTGCGGTAGGTTATAGAGATAAACTGCATTAGCCATTGGT---GTGCGC 775
Db |||:|||||
780 ValLeuPheGlySerGlyTyrHisAsnGlyGlnAlaValSerLeuGlyAlaAlaGly 799

QY 776 TCACGATTACTGATCGCTTTACCGTAAACGGGTGTAGGTTCAATACCTCAATAGGC 835
Db 800 LeuSerAspThrGlyLysSerThrTyrLysIleGlyLeuSerTrpSerAsp---AlaGly 818
QY 836 GGCATGCTTATGCTGCTCTGTTGGTTATGAATTC 871
Db 819 GlyLeuSerGlyGlyValGlyGlySerTyrArgTrp 830

RESULT 13

US-09-540-236-3401
; Sequence 3401, Application US/09540236
; Patent No. 6673910
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATARRHALIS
; FILE REFERENCE: 2709.2005-001
; CURRENT APPLICATION NUMBER: US/09/540.236
; CURRENT FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 3840
; SEQ ID NO 3401
; LENGTH: 878
; TYPE: PRT
; ORGANISM: M.catarrhalis
US-09-540-236-3401

Alignment Scores:
Pred. No.: 0.00239 Length: 878
Score: 107.00 Matches: 42
Percent Similarity: 46.71% Conservative: 29
Best Local Similarity: 27.63% Mismatches: 53
Query Match: 5.22% Indels: 28
DB: 4 Gaps: 8

US-10-030-529A-1 (1-1168) x US-09-540-236-3401 (1-878)

QY 428 CTTGGAATAATCGGACGAGCTGAATTTGATATGATATTTTAAAGTAAAGA 487
Db 750 IleAlaThrAsnLysGlnGluLeuLeuGln-----AsnAsp----- 762
QY 488 CACGATTTAAATATAAAGTTCTGTGACGATTTTCCAAATAAACAATAATTTGAT 547
Db 763 -----ArgLeuAsnArgIleAsnGluThrAsnAsnArgGlnAspGlnLysIleAsp 779
QY 548 ACTATAAGTAATATTACTAGAACTCGGCTCTTATTAGATGATTTCTCGTATGATG 607
Db 780 -----GlnLeuGly-----TyrAlaLeuLys 786
QY 608 GAACAAATACATAATATCAATAGTTGTCTTAA-----GAATGCAAACT-----GGT 658
Db 787 GluGlnGlyGlnHisPheAsnAsnArgIleSerAlaValGluArgGlnThrAlaGlyGly 806
QY 659 TTAGCCAAACCAATCAGCATTTCTATGTTAGTGCACCAAAATGGTGTAGCCAAACGAGC 718
Db 807 IleAlaAsnAlaIleAlaIleAlaThrLeuProSerProSerArgAlaGlyGluHisHis 826
QY 719 GTTTCCTCGGTAGGAGGTTATAGAGATAAACTGCATTAGCCATTGGT---GTCCGC 775
Db 827 ValLeuPheGlySerGlyTyrHisAsnGlyGlnAlaAlaValSerLeuGlyAlaAlaGly 846
QY 776 TCACGCAATTACTGATCGCTTTACCGTAAACGGGTGTAGGTTCAATACCTCAATAGGC 835
Db 847 LeuSerAspThrGlyLysSerThrTyrLysIleGlyLeuSerTrpSerAsp---AlaGly 865
QY 836 GGCATGCTTATGCTGCTCTGTTGGTTATGAATTC 871
Db 866 GlyLeuSerGlyGlyValGlyGlySerTyrArgTrp 877

RESULT 14

US-09-336-447A-9
; Sequence 9, Application US/09336447A
; Patent No. 6310190

GENERAL INFORMATION:

; APPLICANT: HANSEN, ERIC J.
; APPLICANT: ABEI, CHRISTOPH
; APPLICANT: COPE, LESLIE D.
; APPLICANT: MACIVER, ISOBEL
; APPLICANT: FISKE, MICHAEL J.
; APPLICANT: FREDENBURG, ROSS A.
; TITLE OF INVENTION: USPA1 AND USPA2 ANTIGENS OF MORAXELLA CATARRHALIS
; FILE REFERENCE: AMCY:024
; CURRENT APPLICATION NUMBER: US/09/336.447A
; CURRENT FILING DATE: 1999-06-21
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 9
; LENGTH: 941
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-09-336-447A-9

Alignment Scores:

Pred. No.: 0.00245 Length: 941
Score: 107.00 Matches: 41
Percent Similarity: 48.73% Conservative: 36
Best Local Similarity: 25.95% Mismatches: 63
Query Match: 5.22% Indels: 18
DB: 4 Gaps: 8

US-10-030-529A-1 (1-1168) x US-09-336-447A-9 (1-941)

QY 437 AATCGGACGAGCTGAATTTGATATGATATTTTAAACGATTTTAAACACGATTTT 496
Db 788 AsnLysThrAlaIleGluGlnAsnIleAsn-----ArgThrValAlaAsnGlyPhe 804
QY 497 AAATATAAAGTTCTTGTATGATGACGATTTTCCAAATAAACAATAATTT----- 544
Db 805 GluIleGluLysAsnLysAlaGlyIleAlaThrAsnLysGlnGluLeuIleLeuGlnAsn 824
QY 545 GATACATATAAGTAATATTACTAGAACTCGGCTCTTATTAGATGAT----- 592
Db 825 AspArgLeuAsnGln-----IleAsnGluThrAsnAsnHisGlnAspGlnLysIleAspGln 843
QY 593 ---TCTTATGATGATGGAACAAATAATACATAATATCAATAGTTGTCTTAA---GAA 646
Db 844 LeuGlyTyrAlaLeuLysGluGlnGlyGlnHisPheAsnAsnArgIleSerAlaValGlu 863
QY 647 TTCCAAACT-----GTTTAGCCAAACCAATACGATTTGCTATGTTAGTGCACCAAT 700
Db 864 ArgGlnThrAlaGlyGlyIleAlaAsnAlaIleAlaIleAlaThrLeuProSerProSer 883
QY 701 GGTGTAGGCAAAACGAGCGTTTCTGCTCGGTAGGAGGTTATAGAGATAAATCGCATTA 760
Db 884 ArgAlaGlyGluHisValLeuPheGlySerGlyTyrHisAsnGlyGlnAlaAlaVal 903
QY 761 GCCATTGGT---GTCCGCTCACCGATTACTGATCGCTTTACCGCTAAAGCGGTGATCG 817
Db 904 SerLeuGlyAlaAlaGlyLeuSerAspThrGlyLysSerThrTyrLysIleGlyLeuSer 923
QY 818 TTCAATACCTACAAATGCGGCATGCTTATGCTGCTCTCTGTTGTTATGAATTC 871
Db 924 TrpSerAsp---AlaGlyGlyLeuSerGlyGlyValGlyGlySerTyrArgTrp 940

RESULT 15

US-09-268-347-24
; Sequence 24, Application US/09268347
; Patent No. 6335182
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; TITLE OF INVENTION: RECOMBINANT HAEMOPHILUS INFLUENZAE ADHESIN PROTEINS
; FILE REFERENCE: 1038-860
; CURRENT APPLICATION NUMBER: US/09/268.347
; CURRENT FILING DATE: 1999-03-16
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 24
; LENGTH: 1002
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-09-268-347-24

Alignment Scores:
Pred. NO.: 0.00549 Length: 1002
Score: 104.00 Matches: 30
Percent Similarity: 54.22% Conservative: 15
Best Local Similarity: 36.14% Mismatches: 36
Query Match: 5.08% Indels: 2
DB: 4 Gaps: 2

US-10-030-529A-1 (1-1168) x US-09-268-347-24 (1-1002)

QY	626	ATCAATAGTTGTCTAAGAAATTCGCAAACTGGTTAGCCCAACCAATCAGCATTTGTTCTATG	685
DB	921	ValAsnLysValGlyLysArgAlaAspAlaGlyThrAlaSerAlaLeuAlaAlaSerGln	940
QY	686	TTAGTGCACCAACCAATGGTGTAGGCAAAACGAGCGTTTCTGTCGGTAGGAGGTTATAGA	745
DB	941	LeuProGlnAlaSerMetSerGlyLysSerMetValSerIleAlaGlySerSerTyrGln	960
QY	746	GATAAACTGCATTAGCCATTGGTGTGCGCTCAGCATTAATGATCGCTTTACCGCTAAA	805
DB	961	GlyGlnSerGlyLeuAlaIleGlyVal--SerArgIleSerAspAsnGlyLysValIle	979
QY	806	CGGGGTGTAGGTTCAATACCTACAAATGGCGGCATGTCT--TATGGTGTCTTCTGTGGT	862
DB	990	IleArgLeuSerGlyThrThrAsnSerGlnGlyLysThrGlyValAlaAlaGlyValGly	999
QY	863	TATGAATTC	871
DB	1000	TyrGlnTrp	1002

Search completed: May 13, 2004, 08:30:41
Job time : 37 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 15, 2004, 02:41:15 ; Search time 3518 Seconds
(without alignments)
9914.448 Million cell updates/sec

Title: US-10-030-529A-1

Perfect score: 1168

Sequence: 1 ataaatcgcattgacatt.....aagccgttaagtgaggac 1168

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

```

1: em_estba:*
2: em_esthum:*
3: em_esthum:*
4: em_esthum:*
5: em_estov:*
6: em_estov:*
7: em_estro:*
8: em_htc:*
9: gb_estl:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_man:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rtd:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	ID	Description
1	66.8	5.7	1101 29	CNS0039G
2	62	5.3	1200 13	BX437758
3	61.6	5.3	442 29	CNS00H2T
4	59.8	5.1	1151 28	AF416115

C 5	58.8	5.0	747	29	CNS0119Q
C 6	58.8	5.0	1101	29	CNS0039G
C 7	57.2	4.9	1182	13	BX338616
8	56.8	4.9	576	29	CNS035N7
9	56.8	4.9	829	29	CNS011NU
C 10	56.8	4.9	1101	29	CNS00GDR
C 11	56.6	4.8	1165	13	BX338369
C 12	56	4.8	759	29	CNS06QXV
C 13	55.4	4.7	1044	13	BX415231
C 14	55.2	4.7	1101	29	CNS012UN
C 15	55.2	4.7	1101	29	CNS017YG
C 16	55	4.7	950	13	BX415411
C 17	55	4.7	1101	29	CNS012I9
C 18	55	4.7	1201	9	AL524807
C 19	55	4.7	1201	9	AL536104
C 20	54.6	4.7	351	12	BM168097
C 21	54.6	4.7	482	12	BM168129
C 22	54.6	4.7	783	29	CNS00A1S
C 23	54.6	4.7	890	28	AQ026918
C 24	54.6	4.7	1200	13	BX436510
C 25	54.4	4.7	865	28	BH180441
C 26	54.4	4.7	865	29	CNS07MAN
C 27	54	4.6	935	28	B10881
C 28	54	4.6	982	13	BX349562
C 29	54	4.6	989	29	CNS015PZ
C 30	54	4.6	1201	9	AL532464
C 31	53.8	4.6	1073	10	BF274651
C 32	53.8	4.6	1200	13	BX437739
C 33	53.6	4.6	942	29	CNS018GS
C 34	53.4	4.6	1101	29	CNS00EVL
C 35	53.2	4.6	523	29	CE542343
C 36	53.2	4.6	961	14	CK230311
C 37	53	4.5	1101	29	CNS00210
C 38	53	4.5	1201	13	BX402521
C 39	52.8	4.5	834	29	CNS00868
C 40	52.8	4.5	1180	13	BX436369
C 41	52.8	4.5	1225	29	CNS0161D
C 42	52.6	4.5	816	29	CNS02KJV
C 43	52.4	4.5	908	29	AG160668
C 44	52.4	4.5	1101	29	CNS017V2
C 45	52.4	4.5	1200	29	CNS016CO

ALIGNMENTS

RESULT 1
CNS0039G
LOCUS
DEFINITION

CNS0039G 1101 bp DNA linear GSS 03-JUN-1999
Drosophila melanogaster genome survey sequence TET3 end of BAC #
BACR08K10 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

AL063921 GI:4941778

GSS.

Drosophila melanogaster (fruit fly)

Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;

Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 1101)

Genoscope.

Direct Submission

REFERENCE

AUTHORS

TITLE

JOURNAL

Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)

Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).

The BDGP is constructing a physical map of the Drosophila

melanogaster genome using these BACs. For further information

please see <http://www.fruitfly.org> The BDGP Drosophila

melanogaster BAC library was prepared by Kazutoyo Oseegawa and

Aaron Mammosser in Pieter de Jong's laboratory in the Department of


```

506 GTTCTGATGACGATTTTCCAAAATAAACAATAATTTGATCTACTATAAGTAAATATTTA 565
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
690 RKRRATTTTAAAWAWAGAAAGAGAGAAATTTTTTTTATKAGARRGATTTTAAW 749
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
566 CTAGAACTGGGTACTTATTTAGATGATCTTATCGTATGATGGAACAAAATACACATAAT 625
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
750 WTATPARGAGADTTTTTTATATWATTTTTTTTWWAAGATDKAAAAAAMWTTTWTAAAA 809
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
626 ATCAATAA 633
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
810 AAAATTWW 817
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 3
LOCUS      CNS00H2T                      442 bp      DNA      linear      GSS 03-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC:
            BACR35M12 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION  AL074554
VERSION     AL074554.1 GI:4953930
KEYWORDS   GSS.
SOURCE     Drosophila melanogaster (fruit fly)
ORGANISM   Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE  1 (bases 1 to 442)
            Genoscope.
            Direct Submission
            Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seque@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT    Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osogawa and
            Aaron Mamoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain Y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila\_bac.htm.

FEATURES             Location/Qualifiers
     source           1..442
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR35M12"
                     /clone_lib="RPCI-98"
                     /note="end : TET3"

ORIGIN
Query Match      5.3%; Score 61.6; DB 29; Length 442;
Best Local Similarity 46.3%; Pred. No. 0.045;
Matches 130; Conservative 18; Mismatches 133; Indels 0; Gaps 0;

Qy 862 TTATGAATTCATCATTCAGTTTAAATCACTAATCGTTTGGTTTATATAAATAAGGCTAA 921
Db 146 TTWTATTTTATTTTTTTTTTTTATATATATTTTTTTTGAITTTTTTATATATATTTTWTW 205
Qy 922 ATGTTTCTCTCATTTAGCTTCTTATTTATCTTTGTTATAGCTTTTCTGCTTATAA 981
Db 206 WTTTTCNTAAATTTTATTTTTTTTTTTTATTTTTTTTTTTTTTTTTTTTTTTTTTTT 265
Qy 982 AACCGTTTTTTAGCCACTTTTATTAATTAAGCTTATTAAGCTTATCAATCAGTTCTACTT 1041
Db 266 ATWATTTTTTAAWATATTTATTTATATATTTTTTTTTTTTAAATTTTTTTGTTTTTTW 325

506 GTTCTGATGACGATTTTCCAAAATAAACAATAATTTGATCTACTATAAGTAAATATTTA 565
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
690 RKRRATTTTAAAWAWAGAAAGAGAGAAATTTTTTTTATKAGARRGATTTTAAW 749
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
566 CTAGAACTGGGTACTTATTTAGATGATCTTATCGTATGATGGAACAAAATACACATAAT 625
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
750 WTATPARGAGADTTTTTTATATWATTTTTTTTWWAAGATDKAAAAAAMWTTTWTAAAA 809
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
626 ATCAATAA 633
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
810 AAAATTWW 817
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 4
LOCUS      AF416115/c                    1151 bp      DNA      linear      GSS 02-JAN-2002
DEFINITION AF416115 Haemophilus influenzae biotype aegyptius genomic, genomic survey sequence.
ACCESSION  AF416115
VERSION     AF416115.1 GI:18034818
KEYWORDS   GSS.
SOURCE     Haemophilus influenzae biotype aegyptius
ORGANISM   Haemophilus influenzae biotype aegyptius
            Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
            Pasteurellaceae; Haemophilus.
REFERENCE  1 (bases 1 to 1151)
            Smoot, L.M., Franke, D.D., McGillivray, G. and Actis, L.A.
            Genomic analysis of the F3031 Brazilian purpuric fever clone of
            Haemophilus influenzae biogroup aegyptius by PCR-based subtractive
            hybridization
            Infect. Immun. 70 (5), 2694-2699 (2002)
JOURNAL    Microbiology
MEDLINE    21950594
PUBMED     11953414
COMMENT    Contact: Actis LA
            Miami University
            40 Pearson Hall, Oxford, OH 45056, USA
            Email: actisla@muohio.edu
            associated with Brazilian purpuric fever biotype: aegyptius
            isolate = F2
            Class: PCR-based subtractive hybridization.

FEATURES             Location/Qualifiers
     source           1..1151
                     /organism="Haemophilus influenzae biotype aegyptius"
                     /mol_type="genomic DNA"
                     /strain="F3031"
                     /db_xref="taxon:725"
                     /clone_lib="Haemophilus influenzae biotype aegyptius
                     F3031"

ORIGIN
Query Match      5.1%; Score 59.8; DB 28; Length 1151;
Best Local Similarity 63.6%; Pred. No. 0.076;
Matches 91; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

Qy 627 TCAATAAGTTTGTCTAAAGAAATTCAAACTGGTTTGTAGCCAACTCAGCATTTGCTATGT 686
Db 145 TAATAACCTTATAAAGATCTTAAACGTGCTCTGCTGCTCAAGCTGCATTAAATGGTT 86
Qy 687 TAGTCAACCAAAATGGTGTAGGCAAAACGACGGTTTCTGCTCGGTAGGAGGTTATAGAG 746
Db 85 TATTCCAACCGTAAACGTAGGTAAATTAATCTTACTGCTGCTGTAGGTGGTTATAAAT 26
Qy 747 ATAAACTGCATTAGCCATTGGT 769
Db 25 CTCAAACTGCAGTTGCTGTAGGT 3

RESULT 5
LOCUS      CNS011RQ/c                    747 bp      DNA      linear      GSS 26-JUL-1999
DEFINITION Drosophila melanogaster genome survey sequence SP6 end of BAC:
            BACN06A03 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION  AL100640
VERSION     AL100640.1 GI:5612251

506 GTTCTGATGACGATTTTCCAAAATAAACAATAATTTGATCTACTATAAGTAAATATTTA 565
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
690 RKRRATTTTAAAWAWAGAAAGAGAGAAATTTTTTTTATKAGARRGATTTTAAW 749
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
566 CTAGAACTGGGTACTTATTTAGATGATCTTATCGTATGATGGAACAAAATACACATAAT 625
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
750 WTATPARGAGADTTTTTTATATWATTTTTTTTWWAAGATDKAAAAAAMWTTTWTAAAA 809
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
626 ATCAATAA 633
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
810 AAAATTWW 817
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 3
LOCUS      CNS00H2T                      442 bp      DNA      linear      GSS 03-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC:
            BACR35M12 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION  AL074554
VERSION     AL074554.1 GI:4953930
KEYWORDS   GSS.
SOURCE     Drosophila melanogaster (fruit fly)
ORGANISM   Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE  1 (bases 1 to 442)
            Genoscope.
            Direct Submission
            Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seque@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT    Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osogawa and
            Aaron Mamoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain Y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila\_bac.htm.

FEATURES             Location/Qualifiers
     source           1..442
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR35M12"
                     /clone_lib="RPCI-98"
                     /note="end : TET3"

ORIGIN
Query Match      5.3%; Score 61.6; DB 29; Length 442;
Best Local Similarity 46.3%; Pred. No. 0.045;
Matches 130; Conservative 18; Mismatches 133; Indels 0; Gaps 0;

Qy 862 TTATGAATTCATCATTCAGTTTAAATCACTAATCGTTTGGTTTATATAAATAAGGCTAA 921
Db 146 TTWTATTTTATTTTTTTTTTTTATATATATTTTTTTTGAITTTTTTATATATATTTTWTW 205
Qy 922 ATGTTTCTCTCATTTAGCTTCTTATTTATCTTTGTTATAGCTTTTCTGCTTATAA 981
Db 206 WTTTTCNTAAATTTTATTTTTTTTTTTTATTTTTTTTTTTTTTTTTTTTTTTTTTTT 265
Qy 982 AACCGTTTTTTAGCCACTTTTATTAATTAAGCTTATTAAGCTTATCAATCAGTTCTACTT 1041
Db 266 ATWATTTTTTAAWATATTTATTTATATATTTTTTTTTTTTAAATTTTTTTGTTTTTTW 325

```

```

1042 TCACCTTTTTTCCACATATATATCCGCCACTTCTAAAACGGTAAATATTAGTTGGTTAGCC 1101
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
326 TTWTTTTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 385
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1102 TAAATTCGGTACCTTCTATCGGAATTTTTTTTCTAAATGTCTT 1142
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
386 TTWTTTTTTTTTTTTTTTTTTTTATTTTATTTTATTTTATTTT 426
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 4
LOCUS      AF416115/c                    1151 bp      DNA      linear      GSS 02-JAN-2002
DEFINITION AF416115 Haemophilus influenzae biotype aegyptius genomic, genomic survey sequence.
ACCESSION  AF416115
VERSION     AF416115.1 GI:18034818
KEYWORDS   GSS.
SOURCE     Haemophilus influenzae biotype aegyptius
ORGANISM   Haemophilus influenzae biotype aegyptius
            Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
            Pasteurellaceae; Haemophilus.
REFERENCE  1 (bases 1 to 1151)
            Smoot, L.M., Franke, D.D., McGillivray, G. and Actis, L.A.
            Genomic analysis of the F3031 Brazilian purpuric fever clone of
            Haemophilus influenzae biogroup aegyptius by PCR-based subtractive
            hybridization
            Infect. Immun. 70 (5), 2694-2699 (2002)
JOURNAL    Microbiology
MEDLINE    21950594
PUBMED     11953414
COMMENT    Contact: Actis LA
            Miami University
            40 Pearson Hall, Oxford, OH 45056, USA
            Email: actisla@muohio.edu
            associated with Brazilian purpuric fever biotype: aegyptius
            isolate = F2
            Class: PCR-based subtractive hybridization.

FEATURES             Location/Qualifiers
     source           1..1151
                     /organism="Haemophilus influenzae biotype aegyptius"
                     /mol_type="genomic DNA"
                     /strain="F3031"
                     /db_xref="taxon:725"
                     /clone_lib="Haemophilus influenzae biotype aegyptius
                     F3031"

ORIGIN
Query Match      5.1%; Score 59.8; DB 28; Length 1151;
Best Local Similarity 63.6%; Pred. No. 0.076;
Matches 91; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

Qy 627 TCAATAAGTTTGTCTAAAGAAATTCAAACTGGTTTGTAGCCAACTCAGCATTTGCTATGT 686
Db 145 TAATAACCTTATAAAGATCTTAAACGTGCTCTGCTGCTCAAGCTGCATTAAATGGTT 86
Qy 687 TAGTCAACCAAAATGGTGTAGGCAAAACGACGGTTTCTGCTCGGTAGGAGGTTATAGAG 746
Db 85 TATTCCAACCGTAAACGTAGGTAAATTAATCTTACTGCTGCTGTAGGTGGTTATAAAT 26
Qy 747 ATAAACTGCATTAGCCATTGGT 769
Db 25 CTCAAACTGCAGTTGCTGTAGGT 3

RESULT 5
LOCUS      CNS011RQ/c                    747 bp      DNA      linear      GSS 26-JUL-1999
DEFINITION Drosophila melanogaster genome survey sequence SP6 end of BAC:
            BACN06A03 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION  AL100640
VERSION     AL100640.1 GI:5612251

```



```

      :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
722  RDTRWDADADDTAEDDRRRGDDGADAGKGGKTKGRKERRDRATWDRTDAWADAAT 663
      :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
715  GAGCGTTCTGCTGCGGAGGAGGTATAGAGATAAACTGCAATGACCATGGTGTCGG 774
      :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
662  TTDTDUDWDKRRRRKGRRRRTTAAADWWTWKAWDKWMDKTRADRWDRWAADT 603
      :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
775  CTCGCAATTAAGTCTGCTTACCGCTAAAGCGGTGCTAGGTTCAATACCTACAATGG 834
      :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
602  WTDKARDKAWAKARAWRARDRARAARADRRWTGKTTATTTATTTAAARAWAWAWA 543
      :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
835  CGCATGTCCTTATGCTGCTCTGTTGTTATGAATCTTAATCAATFACCTTTAAATCACTAA 894
      :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
542  TTTATTTTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWT 483
      :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
895  TCGTTTGGTTATATAAAAGGCTAATGTTTCTCTCACATAGGCTTCTTATTTA 954
      :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
482  TTTTTTTTTTTTAAWATAAWTATWTWTWTWTWTWTWTWTWTWTWTWTWTWTWT 423
      :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
955  TCTTGTATAGCTTTGCTGCTGTATATAAACCGTTTATTTAGCCACTTTTATTAATTAAGCT 1014
      :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
422  TWTWTWAAAAAAMWWMTWTATATKCCGCCCYCCGCCCYCCCTCCYCCCTCCY 363
      :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
1015  TTTAAGCCTATCAATCAGTCTACTTTCACCTTTTTCACCATATATCC 1064
      :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
362  TCYCCYYYYYYYYYYCCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCT 313

      :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
RESULT 7
BX338616 1162 bp mRNA linear EST 02-MAY-2003
LOCUS
DEFINITION
BX338616 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
clone CSODI061YH17 5-PRIME, mRNA sequence.
ACCESSION
BX338616
VERSION
BX338616.1 GI:30343705
KEYWORDS
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 1162)
Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
Full-length cDNA libraries and normalization
Unpublished (2001)
JOURNAL
COMMENT
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: segre@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. Contact : Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CSODI061CD09QP1.
FEATURES
source
1. 1162
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CSODI061YH17"
/tissue type="PLACENTA COT 25-NORMALIZED"
/clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
/note="1st strand cDNA was primed with a NotI-oligo (dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."

      :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
Query Match 4.9%; Score 57.2; DB 13; Length 1162;
Best Local Similarity 38.3%; Pred. No. 0.24;
Matches 111; Conservative 47; Mismatches 132; Indels 0; Gaps 0;

      :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
808  GGGTGTAGCTCAATACCTACATGCGGCGCATGCTTATGCTCTTCTGTTGTTATGA 867
      :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
853  KKKAAATTTTAAAAAADAAGGGKTTTTTTTTTTGGGTWWTTTTTTTTTTTTTTT 912

```

```

      :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
868  ATTCTAATCATTAGCTTTAATCACTAATCGTTTGGTTATATAATAAAGGCTAATGTTT 927
      :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
913  TTTTWWTTTTTWWAAAAAARAAATTTTTTKKTTTTRRDTTTTTTTTTTTTTTTTTTT 972
      :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
928  CTCCTCACATTAACGCTTTCTTATTTATCTTTGTTATAGCTTTTTCGCTGTTATATAAACCGT 987
      :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
973  TTTTWWTTTTTWWAAGGGRGTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 1032
      :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
988  TTTTTCAGCCTTTTATTAATTAAGCTTTTAAAGCCTATTCATCAGTCTTACTTTCACCTT 1047
      :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
1033  TTTTGGGVAWATWAWAKKTTTTTTTTTTTTTWWAAAAAATAATTTTTTGKKTTTTTT 1092
      :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
1048  TTTTCCCATATATATCCGACATCTCTAAACGGTAATATTAAGTTGGTTT 1097
      :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
1093  TTTTAAADWTAAVAGGYTTKAAAAAADAADWTTTTTTTTTTTTTTTT 1142

      :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
RESULT 8
CNS035N7 576 bp DNA linear GSS 01-SEP-2000
LOCUS
DEFINITION
CNS035N7 Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone
214A06 of library G from Tetraodon nigroviridis, genomic survey
sequence.
ACCESSION
AL228940
AL228940.1 GI:7887933
VERSION
GSS; genome survey sequence.
KEYWORDS
Tetraodon nigroviridis
SOURCE
Tetraodon nigroviridis
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodontidae; Tetraodon.
REFERENCE
1
Roest Crolius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
Saurin,W. and Weissenbach,J.
Estimate of human gene number provided by genome-wide analysis
using Tetraodon nigroviridis DNA sequence
Nat. Genet. 25 (2), 235-238 (2000)
20296633
PUBMED
10835645
REFERENCE
2
Roest Crolius,H., Jaillon,O., Dasilva,C., Ozouf-Costaz,C.,
Fizames,C., Fischer,C., Bouneau,L., Billault,A., Quetier,F.,
Saurin,W., Bernot,A. and Weissenbach,J.
Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis
Genome Res. 10 (7), 939-949 (2000)
20359837
PUBMED
10899143
REFERENCE
3 (bases 1 to 576)
Genoscope.
Direct Submission
Submitted (12-APR-2000) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : segre@genoscope.cns.fr)
- Web : www.genoscope.cns.fr
This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetraodon nigroviridis
genome. For more information, please take a look at
http://www.genoscope.cns.fr/Tetraodon.
FEATURES
source
1. 576
/organism="Tetraodon nigroviridis"
/mol_type="genomic DNA"
/db_xref="taxon:99883"
/clone_lib="G"
/clone="214A06"
/note="Genoscope sequence ID : COAG214BA038P1-end :
PUC-Ori"

      :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
Query Match 4.9%; Score 56.8; DB 29; Length 576;

```



```

Db      987  THMAKHTTTTWTAKTATAGYCDKXVHMHHHTKMMGTMMVCMWMMYNTYHHHTWC 928
Qy      289  GCCAAAGAAGTGGATTTCTAAACAGGCTACTTATCTTGAATTA CAGCAATTATATGCTTA 348
Db      927  ATMCATKTMKTKTTTCTTTTATGTMKTTTTCCTTTTMAKXKATYMMAYKTAKMMWMTMCGM 868
Qy      349  TACTCTCTGTTCTCGTGACATATGCTCCCTGGGCTTCTCCCTAGCCCTACTACTGTTATATCC 408
Db      867  SMKYATMKMHTTMMWMMATMKWTMMTTTTTWTTCMGTGTYAKTCHMMCAITKTCIDGK 808
Qy      409  GATGCTGATCCTGATCAACTTGGAAATAAATCGGAGCAGCTGAAATTTGAATTTGTATAG 468
Db      807  KKSIVAGKAAAMTATTHAHWAHTTTKTAVATHWTKAKYTKYKTWTAMWMMGMKAKATMW 748
Qy      469  TTATTTTAAAGATTAGACACAGATTTTAAATTTAAAGTTCTTCTGATGCACGTATTTCCAA 528
Db      747  AKAYAACWMAATTTTACATWYAGATHAAAWMAWAATTTCTWKAKAATAAKHATTTTGT 688
Qy      529  AAATAAACAATAATGATACATAGTAATATTTTACTAGAACCTGGGTACTTATTTAGA 588
Db      687  TTCCACACATAATWNTTTTTTAWACAAHTTTTTTBAATAAWAATKTAKWATAATWTAYA 628
Qy      589  TGATTTCTTATCGTATGATGAAACAAATAACACATAATATCAATAAGTTGTCTAAAGAAAT 648
Db      627  THACTAKATTTAAKWAATAATTTTTTTTAAKAMATATACHTTTTABWTATAKTKATAMAK 568
Qy      649  GCAA 652
Db      567  TCTA 564

```

RESULT 11
 BX338369/c
 LOCUS
 DEFINITION BX338369 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA clone CSODI058YF14 5-PRIME, mRNA sequence.

ACCESSION BX338369
 VERSION BX338369.1 GI:30343682
 KEYWORDS EST.
 SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1. (bases 1 to 1165)
 Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
 Full-length cDNA libraries and normalization
 Unpublished (2001)
 Contact: Genoscope

Genoscope - Centre National de Sequencage
 BP 191 91006 EVRY cedex - France
 Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr
 Library was constructed by Life Technologies, a division of Invitrogen. Contact : Feng Liang Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com/ Invitrogen Corporation 1600 Faraday Avenue Genoscope sequence ID : CSODI058DC07QPL.

FEATURES

source

```

1. 1165
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CSODI058YF14"
/tissue_type="PLACENTA COT 25-NORMALIZED"
/clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
/note="First strand cDNA was primed with a NotI-oligo (dT) primer. Five prime end enriched, double-strand cDNA was digested with NotI and cloned into the NotI and EcoRV sites of the pCMVSPORT 6 vector. Library was normalized."

```

ORIGIN

Query Match 4.8%; Score 56.6; DB 13; Length 1165;
 Best Local Similarity 37.4%; Pred. No. 0.32; 307; Indels 1; Gaps 1;
 Matches 218; Conservative 57; Mismatches 307; Indels 1; Gaps 1;

```

Qy      469  TTATTTTAAAGATTAGACACAGATTTTAAATTTAAAGTTCTTCTGATGCACGTATTTCCAA 528
Db      1165  WWWTTTTTAAAWMTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 1106
Qy      529  AAATAAACAATAATGATACATAGTAATATTTTACTAGAACCTGGGTACTTATTTAGA 588
Db      1105  TTWTATWATAWMTWMTTTTTTATAWAAWAAWATAATATTTTATTTTATTTTATATATK 1046
Qy      589  TGATTTCTTATCGTATGATGAAACAAATACACATAATATCAATAAGTTGTCTAAAGAAAT 648
Db      1045  TTTTTTTTATWATAWMTTAAAAAAATWAAAAAAATWAAAAAAATWATAATATATAWAA 986
Qy      649  GCAAACTGGTTTAGCCCAACCAATCAGCATTTGTCTATGTTAGTGCACCAATGCTGTAGG 708
Db      985  ATATAATAWMTWMTWAAWAAWATWATAWMTWAAWATAWMTWATAWAAWATAATTAATA 926
Qy      709  CAAAAACGAGCGTTTCTGCTCGGTAGGAGGTATAGAGATAAAACTGCATTTAGCCATTTGG 768
Db      925  TAAAAW-ATWTTTATTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 867
Qy      769  TGTGCGGCTCAGCATTTACTGATCGCTTTACGGCTAAAGCGGTGTAGCGTTCAATACCTA 828
Db      866  AAWAAWATTTTWTWATAWMTAAATWMTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 807
Qy      829  CAATGCGGCGATCTCTATGCTGCTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 888
Db      806  TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 747
Qy      889  CACTAATCGTTTTTGGTTATATAAAGAGCGTAAATGTTTCTCTCCATCAGATTTAGCCTTTCT 948
Db      746  TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 687
Qy      949  TATTATCTTTGTTATAGCTTTTGTGCTGTATAAAGCCGTTTTTTTAGCCACTTTTATTAAT 1008
Db      686  TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 627
Qy      1009  TAAGCTTTTAAAGCTATTCAATCAGTCTACTTCTACTTCTACTTCTACTTCTACTTCTACTT 1051
Db      626  TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 584

```

RESULT 12
 CNS06QXV/c

LOCUS

DEFINITION

Yarrowia lipolytica, genomic survey sequence.

ACCESSION

AL411257

VERSION

AL411257.1 GI:12180512

KEYWORDS

GSS.

SOURCE

Yarrowia lipolytica

Yarrowia lipolytica

Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

1 (bases 1 to 759)

REFERENCE

AUTHORS

Souci, J.L., Aigle, M., Artiguenave, F., Blandin, G.,

Bolotin-Fukuhara, M., Bon, E., Brothier, P., Casaregola, S.,

de-Montigny, J., Dujon, B., Durrens, P., Lepingle, A., Illorente, B.,

Malpertuy, A., Neuveglise, C., Ozier-Kalogeropoulos, O., Potier, S.,

Saurin, W., Tekala, P., Toffano-Nioche, C., Wesolowski-Louvel, M.,

Wincker, P. and Weissenbach, J.

Genomic exploration of the hemiascomycetous yeasts: 1. A set of

yeast species for molecular evolution studies

FEBS Lett. 487 (1), 3-12 (2000)

PUBMED

2 (bases 1 to 759)

REFERENCE

AUTHORS

Casaregola, S., Neuveglise, C., Lepingle, A., Bon, E., Feynrol, C.,

Artiguenave, F., Wincker, P. and Gaillardin, C.

Genomic exploration of the hemiascomycetous yeasts: 17. Yarrowia

lipolytica

FEBS Lett. 487 (1), 95-100 (2000)

20584727

```

PUBMED 11152892
REFERENCE 3 (bases 1 to 759)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage,
2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
secref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT This GSS is part of a random genomic sequencing program of thirteen
yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
5 kb were prepared and both extremities were sequenced. See
keywords for description of this sequence and for the sequence of
the other extremity of this insert.
FEATURES
    source
        Location/Qualifiers
            1..759
                /organism="Yarrowia lipolytica"
                /mol_type="genomic DNA"
                /strain="ClIB 89"
                /db_xref="taxon:4952"
                /clone="AW0AA009H09"
                /clone_lib="AW0AA"
                /note="end : T7"
ORIGIN
Query Match 4.8%; Score 56; DB 29; Length 759;
Best Local Similarity 36.7%; Pred. No. 0.47;
Matches 113; Conservative 50; Mismatches 145; Indels 0; Gaps 0;
QY 842 TCTTATGCTGCTTCGTTGTTATGAATTAATCAATCACTTAATCACTAAATCGTTTT 901
DB 469 TTTTTTTTTTTTTTTGKGKTKTKWTTTTTTTTTTTTTTTTTKTKTKWTTTTTT 410
QY 902 GGTATAATAAAGCGTAAATGTTCTCCTCACATTTAGCGTTTCTTATTTATCTTGT 961
DB 409 TKTTTAAATTTWTTATKGWTTTWTWTTTATKTKWTTTWTGTTT 350
QY 962 TATAGCTTTGCTGTATAAAACCGTTTTTTAGCCACTTTTATTAATTAAGCTTTTAAGC 1021
DB 349 TTTWKTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 290
QY 1022 CTATTCATCAGTCTACTTTCACCTTTTACCATATATCCGCCACTTCTAAACGGT 1081
DB 289 AATTTTGTGTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTATWTTT 230
QY 1082 AATATTAGTTGGTTAGCTAAATGGGTACCTTCTATCGGAATTTTTTCTAAATGTC 1141
DB 229 TTTTWWTATAWATTTATTATWATAGATWTKWATWTTWTKWTTWATWTKWATA 170
QY 1142 TAAATTA 1149
DB 169 TKWATWA 162
RESULT 13
BX415231
LOCUS BX415231 Homo sapiens THYMUS Homo sapiens cDNA clone EST 13-MAY-2003
DEFINITION 3-PRIME, mRNA sequence.
ACCESSION BX415231
VERSION BX415231.1 GI:30637159
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1044)
AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: secref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 5744.r For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0CAP005AF01NP1&cluster=5744.r. Contact :
Peng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ invitrogen Corporation 1600
Paraday Avenue Genoscope sequence ID : CS0CAP005AF01NP1.
Location/Qualifiers
    1..1044
        /organism="Homo sapiens"
        /mol_type="mRNA"
        /db_xref="taxon:9606"
        /clone="CS0CAP005YK01"
        /tissue_type="THYMUS"
        /clone_lib="Homo sapiens THYMUS"
        /note="vector: pCMVSPORT 6; 1st strand cDNA was primed
        with a NotI-oligo(dT) primer. Five prime end enriched,
        double-strand cDNA was digested with Not I and cloned into
        the Not I and EcoRV sites of the pCMVSPORT 6 vector.
        Library was not normalized."
ORIGIN
Query Match 4.7%; Score 55.4; DB 13; Length 1044;
Best Local Similarity 35.1%; Pred. No. 0.56;
Matches 105; Conservative 61; Mismatches 133; Indels 0; Gaps 0;
QY 844 TTATGGTGCTCTGTTGTTATGAATTAATCAATCACTTAATCACTAAATCGTTTGG 903
DB 646 TTTTTHYTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 705
QY 904 TTATAATAAAGCGTAAATGTTCTCCTCACATTTAGCCCTTCTTATTAATCACTTGT 963
DB 706 TTTTMMTHTTTTTTTMMTTTMMTTTMMTTTMMTTTMMTTTMMTTTMMTTTMMTTT 765
QY 964 TAGCTTTTGGTGTATAAAACCGTTTTTTAGCCACTTTTATTAATTAAGCTTTTAAGCT 1023
DB 766 TTTTMMYTTMMHMTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 825
QY 1024 ATTCAATCAGTCTACATTTTCACATTTTTCACCATTTATCGCCACTTCTAAACGGTAA 1083
DB 826 TTTMMTHTTTHTTTTTTTMMTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTMMT 885
QY 1084 TATTAAGTTGTTTAGCTAAATGGGTACCTTCTATCGGAATTTTTTCTAAATGTTCT 1142
DB 886 TTTTWHHTTTTMMTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTMM 944
RESULT 14
CNS012JN/JC 1101 bp DNA linear GSS 26-JUL-1999
LOCUS Drosophila melanogaster genome survey sequence T7 end of BAC
DEFINITION BACN08024 of DrosBAC library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL101645
VERSION AL101645.1 GI:5613256
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : secref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
```

collaboration with the European Drosophila Genome Project (EDGP) - <http://www.edgp.ebi.ac.uk> - This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaud at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBelobAC11.

```

FEATURES
  source
    1. .1101
      /location/Qualifiers
        /organism="Drosophila melanogaster"
        /mol_type="genomic DNA"
        /db_xref="taxon:7227"
        /clone="BACN08024"
        /clone_lib="DrosBAC"
        /plasmid="pBelobAC11"
        /note="end : 17"

ORIGIN
  Query Match      4.7%; Score 55.2; DB 29; Length 1101;
  Best Local Similarity 43.1%; Pred. No. 0.61;
  Matches 132; Conservative 23; Mismatches 151; Indels 0; Gaps 0;

Qy 844 TTATGGTCTCTGTTGTAAGATCTCAATCAATACGTTTAATCACTAATCGTTTGG 903
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 558 TTCTTTTWTWTTCCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 499

Qy 904 TTATAATAAAGCTAAAGTTTCTCTCAATTAAGCTTCTCTCTCTCTCTCTCTCT 963
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 498 TTTATACCTTTTWTTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 439

Qy 964 TAGCTTTTGTCTTATAAAGCGTTTCTTACCGCACTTTTATTAATTAAGCTTTT 1023
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 438 TTGTTTTTYYKATTTTTTWTTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 379

Qy 1024 ATCAATCAGTCTTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1083
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 378 TTTTTCCTTTTTCCTTTTTCCTTTTTCCTTTTTCCTTTTTCCTTTTTCCTTT 319

Qy 1084 TATTAGTCTGTTAGCTTAATGGGTACCTTCTCTCTCTCTCTCTCTCTCTCT 1143
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 318 TKTWTTCCTTTTTCCTTTTTCCTTTTTCCTTTTTCCTTTTTCCTTTTTCCT 259

Qy 1144 AATTA 1149
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 258 ATWTW 253

```

```

RESULT 15
CNS017YG
LOCUS
DEFINITION
  CNS017YG 1101 bp DNA linear GSS 26-JUL-1999
  Drosophila melanogaster genome survey sequence SP6 end of BAC
  BACN37P24 of DrosBAC library from Drosophila melanogaster (fruit
  fly), genomic survey sequence.
ACCESSION
  AL108658
VERSION
  AL108658.1 GI:5628962
KEYWORDS
  GSS.
SOURCE
  Drosophila melanogaster (fruit fly)
  Drosophila melanogaster
  Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;
  Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
  Ephydroidea; Drosophilidae; Drosophila.
  1 (bases 1 to 1101)
REFERENCE
  Genoscope.
  Direct Submission
  Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
  BP 191 91006 EVRY cedex - FRANCE (E-mail : secref@genoscope.cns.fr)
  - Web : www.genoscope.cns.fr
COMMENT
  Determination of this BAC-end sequence was carried out as part of a
  collaboration with the European Drosophila Genome Project (EDGP) -
  http://www.edgp.ebi.ac.uk - This Drosophila melanogaster BAC
  library (Dros BAC) was made by Alain Billaud at CEPH (Centre
  d'Etude du Polymorphisme Humain) with funding provided by a MRC
  project grant. The DNA was prepared from embryos by Alain Bucheton

```

```

FEATURES
  source
    1. .1101
      /location/Qualifiers
        /organism="Drosophila melanogaster"
        /mol_type="genomic DNA"
        /db_xref="taxon:7227"
        /clone="BACN37P24"
        /clone_lib="DrosBAC"
        /plasmid="pBelobAC11"
        /note="end : SP6"

ORIGIN
  Query Match      4.7%; Score 55.2; DB 29; Length 1101;
  Best Local Similarity 38.7%; Pred. No. 0.61;
  Matches 111; Conservative 38; Mismatches 138; Indels 0; Gaps 0;

```

```

Qy 843 CTTATGGTCTCTGTTGTTATGAATCTTAATCATTACGTTTAATCACTAATCGTTT 902
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 189 CAAATWTTCTCTCTATNTATATTTTNNCTTAAATATWTTTNNNTTTT 248

Qy 903 GTTATAATAAAGGCTAAAGTTTCTCTCAATTTAGCCTTCTTATTAATCTTTGTT 962
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 249 WTTTWTTCCTTTTTCCTTTTTCCTTTTTCCTTTTTCCTTTTTCCTTTTTCCT 308

Qy 963 ATAGCTTTGCTCTTATAAAGCGTTTCTTACCGCACTTTTATTAATTAAGCTTT 1022
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 309 WCTACTTTATATATTAATWATWTCCTCATTTTTCCTCATTTTTCCTCATTT 368

Qy 1023 TATTCAATCAGTTCTCTCTTTCCTTTTTCCTCATTTTTCCTCATTTTTCCTCAT 1082
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 369 TATWTTTTCCTTTTTCCTCATTTTTCCTCATTTTTCCTCATTTTTCCTCAT 428

Qy 1083 ATATTAGTTGTTTACGCTTAAATGGGTACCTTCTTATCGGAATTTT 1129
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 429 ATACTAWYCTTTTATWCAACYWTTTACATAAHWCCTCTATHAACCTCT 475

```

Search completed: May 15, 2004, 05:20:56
Job time : 3528 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 15, 2004, 04:22:11 ; Search time 597 Seconds
(without alignments)
8878.461 Million cell updates/sec

Title: US-10-030-529a-1

Perfect score: 1168

Sequence: 1 ataaatcgtcattgacatt.....aagccgttaagtgggac 1168

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2947324 seqs, 2269024515 residues

Total number of hits satisfying chosen parameters: 5894648

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:*

1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
18: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
19: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	57.2	4.9	8139	13	US-10-282-122A-30886
2	54	4.6	1830121	15	Sequence 30886, A
3	54	4.6	1830121	16	Sequence 1, Appli
4	53.2	4.6	2596	10	Sequence 4, Appli
5	53.2	4.6	2673	10	Sequence 12, Appl
6	53.2	4.6	4228	10	Sequence 16, Appl
7	53.2	4.6	94750	13	Sequence 38, Appl
8	52.4	4.5	3295	10	Sequence 8, Appli
9	51.8	4.4	18624	15	Sequence 1675, Ap
10	50.8	4.3	9117	15	Sequence 1041, Ap
11	50.4	4.3	3673778	15	Sequence 1, Appli
12	50	4.3	673	13	Sequence 57885, A
13	49.2	4.2	3673778	15	Sequence 2, Appli
14	49	4.2	18683	15	Sequence 286, App

15	49	4.2	18683	15	US-10-240-452-34	Sequence 34, Appl
16	48.6	4.2	5979	15	US-10-239-676-17	Sequence 17, Appl
17	48.6	4.2	5979	15	US-10-240-453-25	Sequence 25, Appl
18	48	4.1	9964	15	US-10-311-455-71	Sequence 71, Appl
C 19	48	4.1	3673778	15	US-10-312-841-2	Sequence 2, Appli
20	47.6	4.1	14950	15	US-10-311-455-1230	Sequence 2319, Ap
21	47.2	4.0	413	9	US-09-960-352-2919	Sequence 1230, Ap
22	47.2	4.0	431	9	US-09-960-352-5558	Sequence 5558, Ap
23	47.2	4.0	9539	15	US-10-239-676-51	Sequence 51, Appl
24	47.2	4.0	9539	15	US-10-240-453-53	Sequence 53, Appl
C 25	47	4.0	1872	13	US-10-424-599-140661	Sequence 140661,
26	47	4.0	6203	15	US-10-239-676-184	Sequence 184, App
27	47	4.0	6203	15	US-10-240-453-274	Sequence 274, App
28	47	4.0	6757	13	US-10-221-714A-330	Sequence 330, App
29	46.8	4.0	428	9	US-09-960-352-573	Sequence 573, App
30	46.6	4.0	446	9	US-09-960-352-3400	Sequence 3400, Ap
31	46.2	4.0	5989	13	US-10-221-613-172	Sequence 172, App
32	46.2	4.0	5989	15	US-10-240-452-52	Sequence 52, Appl
33	46.2	4.0	6171	15	US-10-311-455-761	Sequence 761, App
34	46.2	4.0	6923	13	US-10-221-613-157	Sequence 157, App
35	46.2	4.0	271990	15	US-10-195-144-87	Sequence 87, Appl
36	46.2	4.0	271990	16	US-10-345-072-87	Sequence 5818, Ap
37	46	3.9	430	9	US-09-960-352-5818	Sequence 168, App
38	46	3.9	5992	13	US-10-221-613-168	Sequence 331, App
39	46	3.9	6664	13	US-10-221-613-331	Sequence 47, Appl
40	46	3.9	9483	13	US-10-221-613-47	Sequence 349, App
41	46	3.9	9483	15	US-10-311-455-349	Sequence 203, App
42	45.8	3.9	5501	13	US-10-221-613-203	Sequence 37, Appl
43	45.8	3.9	5501	15	US-10-204-708-37	Sequence 63, Appl
44	45.8	3.9	5501	17	US-10-240-589C-63	Sequence 2402, Ap
45	45.8	3.9	9110	15	US-10-311-455-2402	

ALIGNMENTS

RESULT 1

US-10-282-122A-30886
; Sequence 30886, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITPA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; PRIOR FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22

```
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 30886
; LENGTH: 8139
; TYPE: DNA
; ORGANISM: Pasteurella multocida
US-10-282-122A-30886

Query Match          4.9%; Score 57.2; DB 13; Length 8139;
Best Local Similarity 51.6%; Pred. No. 0.15;
Matches 131; Conservative 0; Mismatches 123; Indels 0; Gaps 0;

Qy 621 ATATATCAATAAGTGTCTTAAAGAAATGCAAACTGGTTTAGCCCAACCAATCAGCATGT 680
    ||||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 7886 ATCATATCAATAAAGTGGATAAAGATCTCCGTCGGGTATCGCGGTGCCACAGCGGTG 7945

Qy 681 CTATGTAGTCGAACCAATGGTGTAGGCAAAACGAGCGTTTCTGCTCGGTAGGAGTT 740
    ||||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 7946 CGTCTTACACGCGTCCAAATGAAGCAGGTAAGAGCATTTGTTCTCTTGGTGTAGGTTCTT 8005

Qy 741 ATAGAGATAAAACTGCATTAGCCATTGGTGTGGCTCACGCATTACTGCTTTTACCG 800
    ||||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 8006 ATCGTAGTGAATCTGCTATTGTCAGTTGGTTATGCAAGAACTCAGATAACAAATAGATCT 8065

Qy 801 CTAAGCGGGTGTAGCGTTCAATACCTACAATGGCGGCATGCTTTATGGTCTCTCTGTTG 860
    ||||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 8066 CAATCAAACTTGGTGTGTATGAACAGCCGCGTGTGATGTGAACCTTCGTTGGTAGTATCG 8125

Qy 861 GTTATGATTTCTAA 874
    ||||||||| |||||
Db 8126 GTTATCAATGGTAA 8139

RESULT 2
US-10-329-960-1/c
; Sequence 1, Application US/10329960
; Publication No. US20030099277A1
; GENERAL INFORMATION:
; APPLICANT: Fleischmann et al.
; TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus influenzae Rd Genome, Frag
; TITLE OF INVENTION: Thereof, and Uses Thereof
; FILE REFERENCE: PBI86PI
; CURRENT APPLICATION NUMBER: US/10/329,960
; CURRENT FILING DATE: 2003-01-02
; PRIOR APPLICATION NUMBER: US 09/643,990
; PRIOR FILING DATE: 2000-08-23
; PRIOR APPLICATION NUMBER: US 08/487,429
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: US 08/426,787
; PRIOR FILING DATE: 1995-04-21
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1830121
; TYPE: DNA
; ORGANISM: Haemophilus influenzae
; NAME/KEY: misc feature
; LOCATION: (4747)..(4747)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (9921)..(9921)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (10150)..(10150)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (65313)..(65313)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (51805)..(51805)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (55369)..(55369)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (65309)..(65309)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (51786)..(51786)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (51602)..(51602)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (51334)..(51334)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (47036)..(47036)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (45593)..(45593)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (44975)..(44975)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (44905)..(44905)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (44416)..(44416)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (40810)..(40810)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (36636)..(36636)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (36551)..(36551)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (36543)..(36543)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (29298)..(29298)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (10150)..(10150)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
```

```
; NAME/KEY: misc feature
; LOCATION: (80024)..(80024)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (100091)..(100091)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (102696)..(102696)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (105121)..(105121)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (107248)..(107248)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (117136)..(117136)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (119750)..(119750)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (119924)..(119924)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (120038)..(120038)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (121344)..(121344)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (122167)..(122167)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (122336)..(122336)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (131340)..(131340)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (131360)..(131360)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (139910)..(139910)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (140398)..(140398)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (142750)..(142750)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (145058)..(145058)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
```

```
; LOCATION: (145171)..(145171)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (145942)..(145942)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (147197)..(147197)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (150841)..(150841)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (152500)..(152500)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (152530)..(152530)

Query Match      4.6%; Score 54; DB 15; Length 1830121;
Best Local Similarity 61.3%; Pred. No. 8.7;
Matches 87; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

Qy 1024 ATTCAATCAGTTCTACTTTCACCTTTTTCACCATATATATCGGCCACTTCTAAACGGTAA 1083
Db 115616 ATTTAAAGTTTGACGACTTTCGCTGTTTATCATATATATCGGCCACTTCAAGATAGTGA 115557
Qy 1084 TATTAAGTTGGTTTAGCCTAAATTTGGTACCTTCTATCGGAATTTTCTAAATGTTCTA 1143
Db 115556 TTAGCAACCATCAATTCGCAATTCGTGCTTCATCGGTATTTCTCTAAGTGTCTA 115497
Qy 1144 AAATTAAGCGTTTAAAGTGCG 1165
Db 115496 AGATTAACCATTAAGTACG 115475

RESULT 3
US-10-329-670-1/c
; Sequence 1, Application US/10329670
; Publication No. US20040018503A1
; GENERAL INFORMATION:
; APPLICANT: Fleischmann et al.
; TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus influenzae Rd Genome, Frag
; TITLE OF INVENTION: Thereof, and Uses Thereof
; FILE REFERENCE: PB186P1
; CURRENT APPLICATION NUMBER: US/10/329,670
; CURRENT FILING DATE: 2002-12-24
; PRIOR APPLICATION NUMBER: US 09/643,990
; PRIOR FILING DATE: 2000-08-23
; PRIOR APPLICATION NUMBER: US 08/487,429
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: US 08/426,787
; PRIOR FILING DATE: 1995-04-21
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 1
; LENGTH: 1830121
; TYPE: DNA
; ORGANISM: Haemophilus influenzae
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4747)..(4747)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (9921)..(9921)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (10150)..(10150)
```

OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (29238)..(29298)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (36543)..(36543)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (36551)..(36551)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (36836)..(36836)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (40808)..(40810)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (44416)..(44416)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (44905)..(44905)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (44975)..(44975)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (45593)..(45593)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (45732)..(45732)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (47036)..(47036)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (51334)..(51334)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (51602)..(51602)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (51786)..(51786)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (51805)..(51805)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (55369)..(55369)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (65309)..(65309)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (65313)..(65313)
OTHER INFORMATION: n equals a, t, g or c

FEATURE:
NAME/KEY: misc feature
LOCATION: (80024)..(80024)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (100091)..(100091)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (102696)..(102696)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (105121)..(105121)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (107248)..(107248)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (117136)..(117136)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (119750)..(119750)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (119924)..(119924)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (120038)..(120038)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (121344)..(121344)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (122167)..(122167)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (122336)..(122336)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (131340)..(131340)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (131360)..(131360)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (139910)..(139910)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (140398)..(140398)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (142750)..(142750)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (145058)..(145058)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:

```
; NAME/KEY: misc feature
; LOCATION: (145171)..(145171)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (145942)..(145942)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (147197)..(147197)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (150841)..(150841)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (152500)..(152500)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (152530)..(152530)

Query Match      4.6%; Score 54; DB 16; Length 1830121;
Best Local Similarity 61.3%; Pred. No. 8.7;
Matches 87; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

Qy 1024 ATTCAATCAGTTCTACCTTTTTCACCTTTTTCACCATATTATCGGCACCTTCTAAACGGTAA 1083
Db 115616 ATTAAAGTTTGACGACCTTCGCTGTTTATCATATTATTCGCCACCTTCAAGATAGTGA 115557

Qy 1084 TATTAAGTTGGTTAGCTAAATTTGGTACCTTCTATCGGATTTTCTAAATGTTCTA 1143
Db 115556 TTAGCAACCATCAATTTGCGAATTCGTGCTTCATCGGGTATTTCCTCAAGTGTCTA 115497

Qy 1144 AAATTAAGCGCTTAAAGTGCG 1165
Db 115496 AGATTAACCATTAAGTAGG 115475

RESULT 4
US-09-952-267-4
; Sequence 4, Application US/09952267
; Publication No. US20030032772A1
; GENERAL INFORMATION:
; APPLICANT: HANSEN, ERIC J.
; APPLICANT: AEBI, CHRISTOPH
; APPLICANT: COPE, LESLIE D.
; APPLICANT: MACIVER, ISOBEL
; APPLICANT: FISKE, MICHAEL J.
; APPLICANT: FREDENBURG, ROSS A.
; TITLE OF INVENTION: USP1 AND USP2 ANTIGENS OF MORAXELLA CATARRHALIS
; FILE REFERENCE: AMCY.024
; CURRENT APPLICATION NUMBER: US/09/952,267
; PRIOR FILING DATE: 2001-09-12
; PRIOR FILING DATE: 1999-06-21
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 2596
; TYPE: DNA
; ORGANISM: Moraxella catarrhalis
US-09-952-267-4

Query Match      4.6%; Score 53.2; DB 10; Length 2596;
Best Local Similarity 54.0%; Pred. No. 0.6;
Matches 109; Conservative 0; Mismatches 93; Indels 0; Gaps 0;

Qy 625 TATCAATAAGTTGTCTAAAGAAATTCGAACTGTTTAGCCAAACCAATCAGCATTTGCTAT 684
Db 2180 TATCAGAGCTTTAGACAGTAAAGTTGAAACGGTATGGCTGCCACGCTGCACCTTGGTGGCTATGG 2239

Qy 685 GTTAGTGCACCAAAATGGTGTAGGCAAAACAGCGTTTCTGCTGCGGTAGGAGTTATAG 744
Db 2240 TCTATTCCAGCCTTATAGCGTTGTTAAGTTAATGCGACGCTGCACCTTGGTGGCTATGG 2299

Qy 745 AGATAAACTGCATTAGCCATTGGTGTGGCTCAGCATTTACTGATCGCTTTACCGCTAA 804
Db 2300 CTCAAAATCTCGGTTGCTATCGGTGCTATCGTGTGTAATCCAAATCTGGCGTTTAA 2359

Qy 805 AGCGGTTGTAGCGTTCAATACC 826
Db 2360 AGCTGGTGGCGGATTAAATACC 2381

RESULT 6
US-09-952-267-16
; Sequence 16, Application US/09952267
; Publication No. US20030032772A1
; GENERAL INFORMATION:
; APPLICANT: HANSEN, ERIC J.
; APPLICANT: AEBI, CHRISTOPH
; APPLICANT: COPE, LESLIE D.
; APPLICANT: MACIVER, ISOBEL
; APPLICANT: FISKE, MICHAEL J.
; APPLICANT: FREDENBURG, ROSS A.
; TITLE OF INVENTION: USP1 AND USP2 ANTIGENS OF MORAXELLA CATARRHALIS
; FILE REFERENCE: AMCY.024
; CURRENT APPLICATION NUMBER: US/09/952,267
; PRIOR FILING DATE: 2001-09-12
; PRIOR FILING DATE: 1999-06-21
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 2596
; TYPE: DNA
; ORGANISM: Moraxella catarrhalis
US-09-952-267-4

Query Match      4.6%; Score 53.2; DB 10; Length 2596;
Best Local Similarity 54.0%; Pred. No. 0.6;
Matches 109; Conservative 0; Mismatches 93; Indels 0; Gaps 0;

Qy 625 TATCAATAAGTTGTCTAAAGAAATTCGAACTGTTTAGCCAAACCAATCAGCATTTGCTAT 684
Db 2111 TATCAGAGCTTTAGACAGTAAAGTTGAAACGGTATGGCTGCCACGCTGCCTCAAGTGG 2170
```


; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 78
; SEQ ID NO 34
; LENGTH: 18683
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-240-452-34

Query Match 4.2%; Score 49; DB 15; Length 18683;
Best Local Similarity 56.9%; Pred.No.11;
Matches 111; Conservative 0; Mismatches 80; Indels 4; Gaps 1;
Qy 484 AAGACACGATTTAAATTAAGTTCTTGATGCACGTATTTCCAAATAATAACAAATAT 543
Db 3272 ATGAATTTATTTAAATTAAGTTCTTGATGCACGTATTTCCAAATAATAACAAATAT 3327
Qy 544 TGATCTATTAAGTAAATTTTACTAGAACTGGGTACTTTATTTAGATGATCTTATCGTAT 603
Db 3328 TAAAGACATATTAATAGTATGGTAGAGAGATTTGTAATTTATATTTGATAAAGAT 3387
Qy 604 GATGGAACAAATACACATATATCAATAAGTTGCTCTAAAGAAATTGCAAACTGGTTTACG 663
Db 3388 TAATATGTAATAATATAAAGAAATATAAATAATTTAATAGAGGAAATAAATTTGATTTAGA 3447
Qy 664 CAACCAATCAGCAT 678
Db 3448 AATAAGTAAAGGATT 3462

Search completed: May 15, 2004, 06:52:15
Job time : 637 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 15, 2004, 02:44:00 ; Search time 112 Seconds
(without alignments)
5787.345 Million cell updates/sec

Title: US-10-030-529A-1
Perfect score: 1168
Sequence: 1 ataatacgtcattgacatt.....aagcgttaaaggtggac 1168

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA: *
1: /cgn2_6/ptodata/2/ina/5A COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PTUS COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	57.6	4.9	7218	1	US-08-232-463-14
2	54	4.6	1830121	4	US-09-557-884-1
3	54	4.6	1830121	4	US-09-643-990A-1
4	53.2	4.6	2596	4	US-09-336-447A-4
5	53.2	4.6	2604	4	US-09-540-236-756
6	53.2	4.6	2673	4	US-09-336-447A-12
7	53.2	4.6	4228	4	US-09-336-447A-16
8	53.2	4.6	94750	4	US-09-596-002-38
9	52.4	4.5	3295	4	US-09-336-447A-8
10	45.8	3.9	5501	4	US-10-204-708-37
11	45.4	3.9	640681	4	US-09-790-988-1
12	45	3.9	640681	4	US-09-790-988-1
13	44	3.8	11049	4	US-10-204-708-22
14	43.4	3.7	2037	3	US-08-913-942-14
15	43.4	3.7	2079	4	US-09-268-347-25
16	43.4	3.7	3294	1	US-08-409-995-1
17	43.4	3.7	3294	3	US-08-685-467-1
18	43.4	3.7	3294	3	US-08-913-942-1
19	43.4	3.7	3294	4	US-09-268-347-43
20	43.4	3.7	3300	4	US-09-268-347-31
21	43.4	3.7	7253	4	US-09-268-347-35
22	43.4	3.7	7291	3	US-08-913-942-3
23	43.2	3.7	3477	4	US-09-462-720-3
24	43.2	3.7	6243	2	US-09-056-075-1
25	42.8	3.7	1311	4	US-09-489-039A-4614
26	42.6	3.6	3030	4	US-09-268-347-29
27	42.6	3.6	3036	4	US-09-268-347-23

28	42.4	3.6	10467	4	US-10-204-708-2	Sequence 2, Appli
29	42.2	3.6	6866	4	US-10-204-708-20	Sequence 20, Appl
30	42.2	3.6	19233	4	US-10-204-708-45	Sequence 45, Appl
31	42	3.6	1122	4	US-09-543-681A-3292	Sequence 3292, Ap
32	41.8	3.6	837	3	US-08-998-416-288	Sequence 288, App
33	41.6	3.6	11049	4	US-10-204-708-23	Sequence 23, Appl
34	41.4	3.5	5562	4	US-10-204-708-63	Sequence 63, Appl
35	41.4	3.5	19124	2	US-08-487-826B-13	Sequence 13, Appl
36	41	3.5	1575	4	US-09-543-681A-615	Sequence 615, App
37	41	3.5	1812	4	US-09-268-347-37	Sequence 37, Appl
38	41	3.5	3033	4	US-09-134-001C-2341	Sequence 2341, Ap
39	40.6	3.5	636	3	US-08-998-416-1137	Sequence 1137, Ap
40	40.4	3.4	1664976	4	US-08-916-421B-1	Sequence 1, Appli
41	40.2	3.4	14066	4	US-09-601-198-56	Sequence 56, Appl
42	40	3.4	5852	1	US-07-867-106-2	Sequence 2, Appli
43	40	3.4	6306	4	US-10-204-708-49	Sequence 49, Appl
44	39.8	3.4	615	3	US-08-998-416-186	Sequence 186, App
45	39.8	3.4	636	3	US-08-998-416-1137	Sequence 1137, Ap

ALIGNMENTS

RESULT 1
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: ptzgt-Fls
US-08-232-463-14
Query Match 4.9%; Score 57.6; DB 1; Length 7218;

Best Local Similarity 15.6%; Pred. No. 0.00019;
Matches 81; Conservative 199; Mismatches 238; Indels 0; Gaps 0;

Qy 556 TAAATATTACTAGACCTGGGTACTTATTTAGATGATCTTTATCGTATGATGGAACAAA 615
Db 876 TAAAGCAATTCATACACACAGCAGTTAGTTTACCAACATTCAGATAAGTAATAGAT 935

Qy 616 TACATAATATCAATAGTTGTCTAAAGAATTCAAACCTGGTTTAGCCAACCAATCAGC 675
Db 936 TCAAAATATTATAAAGCGTTTACGTTGAAATGTCATCGAGTCGCGCTACTATAACTA 995

Qy 676 ATTGCTATGTTAGTGACCAAAATGGTGTAGGCAAAAGAGCGCTTTCTGTCGGTAGG 735
Db 996 TTTTTCCTTCGTTGGCCATAGCCTCACAGAATTAATCCGAGCTTGGCTGAGGTCGAG 1055

Qy 736 AGGTTATAGAGATAAACTGCAATAGCATTGGTGTGGCTCACGCATTAAGTATCGCTT 795
Db 1056 GAGCTGCGATYYY 1115

Qy 796 TACGCTAAAGCGGTGTAGCGTTCAATACCTACAAATGCGCGCATGCTTTATGTCGCTC 855
Db 1116 YYY 1175

Qy 856 TGTGTTATGAATCTAATCAATCAATGAGTTTAACTCAATCAATGCTTTGTTTATAATAAAA 915
Db 1176 YYY 1235

Qy 916 GGCTAAATGTTCTCCACATTTAGCCTTCTTATTTATCTTTGTTATAGCTTTTCTG 975
Db 1236 YYY 1295

Qy 976 TTATAAAACCGTTTTTAGCCACTTTTATTAATTAAGCTTTTAAAGCCTTATCAATCAGTT 1035
Db 1296 YYY 1355

Qy 1036 CTACTTTCACCTTTTTCACATATATATCCGCCACTTCT 1073
Db 1356 YYY 1393

RESULT 2
US-09-557-884-1/c
; Sequence 1, Application US/09557884
; Patent No. 6506581
; GENERAL INFORMATION:
; APPLICANT: Fleischmann et al.
; TITLE OF INVENTION: The Nucleotide sequence of
; the Haemophilus influenzae Rd Genome, Fragments
; Thereof, and Uses Thereof
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: MD
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3 1/2 inch diskette
; COMPUTER: Dell Pentium
; OPERATING SYSTEM: MS DOS v6.22
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/557,884
; FILING DATE: 25-Apr-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/476,102
; FILING DATE: JUN-5-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Michelle S. Marks
; REGISTRATION NUMBER: 41,971
; REFERENCE/DOCKET NUMBER: PB186P3

TELECOMMUNICATION INFORMATION:
TELEPHONE: 301-309-8504
TELEFAX: 301-309-8439
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1830121 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-557-884-1

Query Match 4.6%; Score 54; DB 4; Length 1830121;
Best Local Similarity 61.3%; Pred. No. 0.0066;
Matches 87; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

Qy 1024 ATTCAATCACTTTCTACCTTTTCCACCTTTTTCACCATATATCGCCACTTCTAAACGGTAA 1083
Db 115616 ATTAAAGTTTGACGACTTCGCTGTTTATCATATATATCGCCACTTCAAGAATAGTGA 115557

Qy 1084 TATTAAGTTGTTAGCCTAAATTTGGGTACCTTCTATCGAAATTTTTTCTAAATGTTCTA 1143
Db 115556 TTAGCAACCATCAATTCGCAATCGTGCCTTCATCGGTATTTCTCTAAGTGTCTA 115497

Qy 1144 AAATTAAGCGGTTAAAGTGCG 1165
Db 115496 AGATTAACCATTAAGTAGC 115475

RESULT 3
US-09-643-990A-1/c
; Sequence 1, Application US/09643990A
; Patent No. 6528289
; GENERAL INFORMATION:
; APPLICANT: Robert D. Fleischmann
; Mark D. Adams
; Owen White
; Hamilton O. Smith
; J. Craig Venter
; TITLE OF INVENTION: The Nucleotide sequence of
; the Haemophilus influenzae Rd Genome, Fragments
; Thereof, and Uses Thereof
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville,
; STATE: MD
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3 1/2 inch diskette
; COMPUTER: Dell Pentium
; OPERATING SYSTEM: MS DOS v6.22
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/643,990A
; FILING DATE: 23-Aug-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/487,429
; FILING DATE: 1995-06-07
; APPLICATION NUMBER: 08/426,787
; FILING DATE: 1995-04-21
; ATTORNEY/AGENT INFORMATION:
; NAME: Kenley K. Hoover
; REGISTRATION NUMBER: 40,302
; REFERENCE/DOCKET NUMBER: PB186P1C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 301-610-5790
; TELEFAX: 310-309-8439
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:

```
;
; LENGTH: 1830121 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-643-990A-1

Query Match      4.6%; Score 54; DB 4; Length 1830121;
Best Local Similarity 61.3%; Pred. No. 0.0066;
Matches 87; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY 1024 ATTCAATCAGTCTTACCTTTTTCACCTTTTTCACCAATATTATCGGCACCTCTTAAACCGGTAA 1083
DB 115616 ATTAAAGTTTGACGACTTTCGCTGTTTATCATATTATCGCCACCTTCAAGATAGTA 115557
QY 1084 TATTAAGTTGGTTAGCCTAAATGGGTACCTTCTATCGGAATTTTTCCTAAATGTTCTA 1143
DB 115556 TTAGCAAAACCATCAATTTGCAAAATCGTCTTCATCGGTATTTCTCTAAGTGTTCTA 115497
QY 1144 AATTAAGCGGTTAAAGTGGG 1165
DB 115496 AGATTAAACCATTAAGAGTAGC 115475

RESULT 4
US-09-336-447A-4
; Sequence 4, Application US/09336447A
; Patent No. 6310190
; GENERAL INFORMATION:
; APPLICANT: HANSEN, ERIC J.
; APPLICANT: AEBI, CHRISTOPH
; APPLICANT: COPE, LESLIE D.
; APPLICANT: MACIVER, ISOBEL
; APPLICANT: FISKE, MICHAEL J.
; APPLICANT: FREDENBURG, ROSS A.
; TITLE OF INVENTION: USP1 AND USP2 ANTIGENS OF MORAXELLA CATARRHALIS
; FILE REFERENCE: AMCY.024
; CURRENT APPLICATION NUMBER: US/09/336,447A
; CURRENT FILING DATE: 1999-06-21
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 4
; LENGTH: 2596
; TYPE: DNA
; ORGANISM: Moraxella catarrhalis
; US-09-336-447A-4

Query Match      4.6%; Score 53.2; DB 4; Length 2596;
Best Local Similarity 54.0%; Pred. No. 0.0016;
Matches 109; Conservative 0; Mismatches 93; Indels 0; Gaps 0;

QY 625 TATCAATAAGTTGCTTAAAGAATTGCAAACTGGTTTAGCCAAACCAATCAGCATTTGTCTAT 684
DB 2111 TATCAGAGCTTTAGACAGTAAAGTTGAAACGGTATGCTGCGCCCAAGCTGCCCTAAGTGG 2170
QY 685 GTTAGTGCACCAAAATGGTTAGCAAAACAGCGTTTCTGCTGCGGTAGAGGTTATAG 744
DB 2171 TCTATTCCAGCTTTATAGCGTTGGTAAGTTTAATGCGACCGCTGCACCTTGGTGGCTATGG 2230
QY 745 AGATAAACTGCATTAGCCATTGGTGTGCGCTACGCAATTAATGATCGCTTTACCGGTAA 804
DB 2231 CTCAAAATCTGCGGTTGCTATCGGTGCTGCTATCGTATCGTGAATCCAAATCTGCGGTTAA 2290
QY 805 AGCGGGTGTAGCGTTCAATACC 826
DB 2291 AGCTGGTGGCGGATTAAATACC 2312

RESULT 5
US-09-540-236-756
; Sequence 756, Application US/09540236
; Patent No. 6673910
; GENERAL INFORMATION:
; APPLICANT: GARY L. BRETON ET AL.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATARRHALIS
; FILE REFERENCE: 2709.2005-001
; CURRENT APPLICATION NUMBER: US/09/540,236
; CURRENT FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 3840
; SEQ ID NO 756
; LENGTH: 2604
; TYPE: DNA
; ORGANISM: M. catarrhalis
; US-09-540-236-756

Query Match      4.6%; Score 53.2; DB 4; Length 2604;
Best Local Similarity 54.0%; Pred. No. 0.0016;
Matches 109; Conservative 0; Mismatches 93; Indels 0; Gaps 0;

QY 625 TATCAATAAGTTGCTTAAAGAATTGCAAACTGGTTTAGCCAAACCAATCAGCATTTGTCTAT 684
DB 2352 TATCAGAGCTTTAGACAGTAAAGTTGAAACGGTATGCTGCGCCCAAGCTGCCCTAAGTGG 2411
QY 685 GTTAGTGCACCAAAATGGTTAGGCAAAACAGCGTTTCTGCTGCGGTAGAGGTTATAG 744
DB 2412 TCTATTCCAGCTTTATAGCGTTTGGTAAGTTTAATGCGACCGCTGCACCTTGGTGGCTATGG 2471
QY 745 AGATAAACTGCATTAGCCATTGGTGTGCGCTCAGCATTAATGATCGCTTTACCGGTAA 804
DB 2472 CTCAAAATCTGCGGTTGCTATCGGTGCTGCTATCGTGTGAATCCAAATCTGCGGTTTAA 2531
QY 805 AGCGGGTGTAGCGTTCAATACC 826
DB 2532 AGCTGGTGGCGGATTAAATACC 2553

RESULT 6
US-09-336-447A-12
; Sequence 12, Application US/09336447A
; Patent No. 6310190
; GENERAL INFORMATION:
; APPLICANT: HANSEN, ERIC J.
; APPLICANT: AEBI, CHRISTOPH
; APPLICANT: COPE, LESLIE D.
; APPLICANT: MACIVER, ISOBEL
; APPLICANT: FISKE, MICHAEL J.
; APPLICANT: FREDENBURG, ROSS A.
; TITLE OF INVENTION: USP1 AND USP2 ANTIGENS OF MORAXELLA CATARRHALIS
; FILE REFERENCE: AMCY.024
; CURRENT APPLICATION NUMBER: US/09/336,447A
; CURRENT FILING DATE: 1999-06-21
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 12
; LENGTH: 2673
; TYPE: DNA
; ORGANISM: Moraxella catarrhalis
; US-09-336-447A-12

Query Match      4.6%; Score 53.2; DB 4; Length 2673;
Best Local Similarity 54.0%; Pred. No. 0.0016;
Matches 109; Conservative 0; Mismatches 93; Indels 0; Gaps 0;

QY 625 TATCAATAAGTTGCTTAAAGAATTGCAAACTGGTTTAGCCAAACCAATCAGCATTTGTCTAT 684
DB 2180 TATCAGAGCTTTAGACAGTAAAGTTGAAACGGTATGCTGCGCCCAAGCTGCCCTAAGTGG 2239
QY 685 GTTAGTGCACCAAAATGGTTAGGCAAAACAGCGTTTCTGCTGCGGTAGAGGTTATAG 744
DB 2240 TCTATTCCAGCTTTATAGCGTTTGGTAAGTTTAATGCGACCGCTGCACCTTGGTGGCTATGG 2299
QY 745 AGATAAACTGCATTAGCCATTGGTGTGCGCTCAGCATTAATGATCGCTTTACCGGTAA 804
DB 2300 CTCAAAATCTGCGGTTGCTATCGGTGCTGCTATCGTGTGAATCCAAATCTGCGGTTTAA 2359
```

```
Qy 805 AGCGGGTGTAGCGTTCAATACC 826
Db 2360 AGCTGGTGGCGCGATTAAATACC 2381

RESULT 7
US-09-336-447A-16
; Sequence 16, Application US/09336447A
; Patent No. 6310190
; GENERAL INFORMATION:
; APPLICANT: HANSEN, ERIC J.
; APPLICANT: AEBI, CHRISTOPH
; APPLICANT: COPE, LESLIE D.
; APPLICANT: MACIVER, ISOBEL
; APPLICANT: FISKE, MICHAEL J.
; APPLICANT: FREDENBURG, ROSS A.
; TITLE OF INVENTION: USP1 AND USP2 ANTIGENS OF MORAXELLA CATARRHALIS
; FILE REFERENCE: AMCY:024
; CURRENT APPLICATION NUMBER: US/09/336.447A
; CURRENT FILING DATE: 1999-06-21
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 16
; LENGTH: 4228
; TYPE: DNA
; ORGANISM: Moraxella catarrhalis
US-09-336-447A-16

Query Match 4.6%; Score 53.2; DB 4; Length 4228;
Best Local Similarity 54.0%; Pred. No. 0.0018;
Matches 109; Conservative 0; Mismatches 93; Indels 0; Gaps 0;

Qy 625 TATCAATAAGTGTCTAAAGAAATGCAAACTGTTTAGCCAAACCAATCAGCATGTCTAT 684
Db 3629 TATCAGAGCTTTAGACAGTAAGTTGAAACGGTATGCTGCCCAAGTGCCTTAAGTGG 3688

Qy 685 GTTAGTCAACCAAAATGGTGTAGGCAAAACAGAGGTTTCTGCTCGGTAGGAGTTATAG 744
Db 3689 TCTATTCCAGCTTTATAGCGTTGGTAAAGTTTAAATGCGACCGCTGCACTTGGTGGCTATGG 3748

Qy 745 AGATAAACTCGATTAGCCATTGCTGCGGTCAACCAATCTAGCATCGCTTTACCGTTAA 804
Db 3749 CTCAAAATCTCGGTTGCTATCGGTGCTGCTATCGTGAATCCAAATCTGGCGGTTAA 3808

Qy 805 AGCGGGTGTAGCGTTCAATACC 826
Db 3809 AGCTGGTGGCGCGATTAAATACC 3830

RESULT 8
US-09-596-002-38
; Sequence 38, Application US/09596002
; Patent No. 6632636
; GENERAL INFORMATION:
; APPLICANT: Lagace, Robert, E.
; APPLICANT: Patterson, Chandra
; APPLICANT: Berg, Kim, L.
; APPLICANT: BEIG, Kim, L.
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCES OF MORAXELLA CATARRHALIS GENOME
; FILE REFERENCE: PM-0008-4 US
; CURRENT APPLICATION NUMBER: US/09/596.002
; CURRENT FILING DATE: 2000-05-16
; PRIOR APPLICATION NUMBER: 60/140,121
; PRIOR FILING DATE: 1999-06-18
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PERL Program
; SEQ ID NO 38
; LENGTH: 94750
; TYPE: DNA
; ORGANISM: M. catarrhalis
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte template ID No. 6632636 38
; PUBLICATION INFORMATION:

US-09-596-002-38
Query Match 4.6%; Score 53.2; DB 4; Length 94750;
Best Local Similarity 54.0%; Pred. No. 0.0044;
Matches 109; Conservative 0; Mismatches 93; Indels 0; Gaps 0;

Qy 625 TATCAATAAGTGTCTAAAGAAATGCAAACTGTTTAGCCAAACCAATCAGCATGTCTAT 684
Db 26785 TATCAGAGCTTTAGACAGTAAGTTGAAACGGTATGCTGCCCAAGTGCCTTAAGTGG 26844

Qy 685 GTTAGTCAACCAAAATGGTGTAGGCAAAACAGAGGTTTCTGCTCGGTAGGAGTTATAG 744
Db 26845 TCTATTCCAGCTTTATAGCGTTGGTAAAGTTTAAATGCGACCGCTGCACTTGGTGGCTATGG 26904

Qy 745 AGATAAACTGCATTAGCCATTGCTGCGGTCAACCAATCTAGCATCGCTTTACCGCTAA 804
Db 26905 CTCAAAATCTGCGGTGCTATCGGTGCTGCTATCGTGAATCCAAATCTGGCGGTTAA 26964

Qy 805 AGCGGGTGTAGCGTTCAATACC 826
Db 26965 AGCTGGTGGCGGATTAAATACC 26986

RESULT 9
US-09-336-447A-8
; Sequence 8, Application US/09336447A
; Patent No. 6310190
; GENERAL INFORMATION:
; APPLICANT: HANSEN, ERIC J.
; APPLICANT: AEBI, CHRISTOPH
; APPLICANT: COPE, LESLIE D.
; APPLICANT: MACIVER, ISOBEL
; APPLICANT: FISKE, MICHAEL J.
; APPLICANT: FREDENBURG, ROSS A.
; TITLE OF INVENTION: USP1 AND USP2 ANTIGENS OF MORAXELLA CATARRHALIS
; FILE REFERENCE: AMCY:024
; CURRENT APPLICATION NUMBER: US/09/336.447A
; CURRENT FILING DATE: 1999-06-21
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 3295
; TYPE: DNA
; ORGANISM: Moraxella catarrhalis
US-09-336-447A-8

Query Match 4.5%; Score 52.4; DB 4; Length 3295;
Best Local Similarity 55.5%; Pred. No. 0.0026;
Matches 101; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

Qy 645 AATTGCAAACTGGTTTAGCCAAACCAATCAGCATGTCTATGTTAGTCAACCAATAGTG 704
Db 2854 AGTTGAAAACGGTATGGCTGCCCAAGCTGCCCTAAGTGGTCTATTCAGGCTTATAGCG 2913

Qy 705 TAGGCAAAACGAGCGTTTCTGCTGCGGTAGGAGTTATAGAGATAAAAACCTGCATTAGCCA 764
Db 2914 TTGGTAAGTTTAAATGCGACCGCTGCACTTGGTGGCTATGGCTCAAAATCTCGGTTGCTA 2973

Qy 765 TTGGTGTGCGCTCAGCATTTACTGATCGCTTTACCGGTAAAGCGGGTGTAGCGTTCAATA 824
Db 2974 TCGGTGCTGGCTATCGTGTGAATCCAAATCTGGCGGTTTAAAGCTGGTGGCGGTAATA 3033

Qy 825 CC 826
Db 3034 CC 3035

RESULT 10
US-10-204-708-37
; Sequence 37, Application US/10204708
; Patent No. 6677731
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
```

```

; LENGTH: 640681
; TYPE: DNA
; ORGANISM: Buchnera sp.
US-09-790-988-1

Query Match      3.9%; Score 45.4; DB 4; Length 640681;
Best Local Similarity 48.0%; Pred. No. 0.53;
Matches 130; Conservative 0; Mismatches 141; Indels 0; Gaps 0;

Qy 458 AATTGTTATAGTTATTTTAAACGATTTTAAGACACGATTTTAAATTAAGAGTTCTTGTATGCA 517
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 209572 AATTTTAAAAATATATGAGGACACTGATAAATTTTATTAAAAATTAAGAACTAGACGTA 209513

Qy 518 CGTATTTTCCAAAAATAACAAAAATTCATACTATAAGTAATAATTTTACTAGAACTGGGT 577
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 209512 TATTTTTCAAATAAAATCAAAACAGCGCGCTTAATGTAGGTAAATGTATTACGATTTGTAGCA 209453

Qy 578 ACTTATTTAGATGATTCCTATCGTGTATGATGAAACAAAAATACACATAATATCAATAAGTTG 637
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 209452 ACAATAGACAAAAAAGACAATTTTATAAACAATTGAAGAAGATTAAATTAATAATCA 209393

Qy 638 TCTAAAGAATTGCAAACTGGTTTAGCCCAACCAATCAGCATTTGTCTATGTTAGTGCAACCA 697
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 209392 TTATACAAAGTAAAAAATGGTGAATGCACTAGCATTTTATACTAATTTATTATCAACCA 209333

Qy 698 AATGGTGTAGCCAAAAACGAGCGTTTCTGCTG 728
      ||||| ||||| ||||| ||||| |||||
Db 209332 ATTCCTCTGGTACTAAGAGGCTACGGTGTG 209302

RESULT 12
US-09-790-988-1
; Sequence 1, Application US/09790988
; Patent No. 6632935
; GENERAL INFORMATION:
; APPLICANT: SHIGENOBU, SHUJI
; APPLICANT: WATANABE, HIDEMI
; APPLICANT: HATTORI, MASAHIRA
; APPLICANT: SAKAKI, YOSHIYUKI
; TITLE OF INVENTION: GENOME DNA OF BACTERIAL SYMBIONT OF APHIDS
; FILE REFERENCE: 081356/0159
; CURRENT APPLICATION NUMBER: US/09/790,988
; CURRENT FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: JP2000-107160
; PRIOR FILING DATE: 2000-04-07
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 640681
; TYPE: DNA
; ORGANISM: Buchnera sp.
US-09-790-988-1

Query Match      3.9%; Score 45; DB 4; Length 640681;
Best Local Similarity 47.8%; Pred. No. 0.66;
Matches 162; Conservative 0; Mismatches 175; Indels 2; Gaps 1;

Qy 422 GATCAACTTCGAATAAAATCGGCAGCAGCTGAAATTTGAATTTGTATAGTTATTTTAAACGAT 481
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 165491 GATGCCAATCTGAAGAGTTCAAAAAGAAAGCAAAAGAATTAATATATCAATCTTTCAAC 165508

Qy 482 TTAAGACACGATTTTAAATTAATAAGTTCTTCGATGCACGTATTTCCAAAAATAAACAATAAT 541
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 165551 TTCAGAAAAAATTT--AATAATATATTAATAATATAAATTTCTTTAATAATTAATCTGATT 165608

Qy 542 ATTGATACTATAAGTAATAATTTTACTAGAACTGGGTACTTAATTAGATGATTTCTTTATCGT 601
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 165609 TTTTAGATGATAATATTATTATATATTTTAAACTGAGTGGTGAGAGATATATGATTGATTTA 165668

Qy 602 ATGATGAAACAAAATACACATAATCAATAAGTTGTCTTAAGANTTGCNAACTGGTTTA 661
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 165669 ATGCCAAAAAATAATACATTTTATTAAAAAATAAATAAGAAAGACTCAATAATAATCA 165728

```

QY 662 GCCAACCAATCAGCATTTCTATGTTAGTGCACCAAAATGGTGTAGGCAAAACGAGCGTT 721
 Db 165729 GTAATTAATAATTTATTTGATTTCTTTGGTAAAGAATTTGTTTTCATCAAAATTACCTTA 165788
 QY 722 TCTGTCGGGTAGGAGTTATAGAGATAAAACATGCATTA 760
 Db 165789 ACCGGTTTTCTATATTTTGGATTGATAAGACATTATTA 165827

RESULT 13
 US-10-204-708-22
 ; Sequence 22, Application US/10204708
 ; Patent No. 667731
 ; GENERAL INFORMATION:
 ; APPLICANT: OLEK, Alexander
 ; APPLICANT: PIEPENBROCK, Christian
 ; APPLICANT: BERLIN, Kurt
 ; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
 ; TITLE OF INVENTION: by Assessing DNA Methylation
 ; FILE REFERENCE: 5013.1012
 ; CURRENT APPLICATION NUMBER: US/10/204,708
 ; PRIOR FILING DATE: 2003-05-06
 ; PRIOR APPLICATION NUMBER: PCT/EP01/03971
 ; PRIOR FILING DATE: 2001-04-06
 ; PRIOR APPLICATION NUMBER: DE 10019058.8
 ; PRIOR FILING DATE: 2000-04-06
 ; PRIOR APPLICATION NUMBER: DE 10019173.8
 ; PRIOR FILING DATE: 2000-04-07
 ; PRIOR APPLICATION NUMBER: DE 10032529.7
 ; PRIOR FILING DATE: 2000-06-30
 ; PRIOR APPLICATION NUMBER: DE 10043826.1
 ; PRIOR FILING DATE: 2000-09-01
 ; NUMBER OF SEQ ID NOS: 98
 ; SEQ ID NO 22
 ; LENGTH: 11049
 ; TYPE: DNA
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
 US-10-204-708-22

Query Match 3.8%; Score 44; DB 4; Length 11049;
 Best Local Similarity 51.5%; Pred. No. 0.36;
 Matches 101; Conservative 0; Mismatches 95; Indels 0; Gaps 0;
 QY 822 ATACCTACATCGCGCATGCTCTATGCTGCTCTGTTGTTATGATTAATCTAATCATTA 881
 Db 1402 ATGATTAATAATGAATTTTATGTTTATTTATTTATTTATTTATTTATTTATTTATTT 1461
 QY 882 GTTTAATCACTAATCGTTTGGTTATAATAAAAGGCTAAATGTTTCTCTCACAATTTAG 941
 Db 1462 ATGTTTTTGTATTTATTTAGTAGTAATGTTGGGTATTTATTTATTTATTTATTTATTTAG 1521
 QY 942 CCTTCTTATTAATCTGTTTATGCTTTTGTGTTTATATAAAACCGTTTTTTTATGCCACTTT 1001
 Db 1522 GTTTATTAATTTTATTTTAAATGTTTATTTGAAGGTATATATTAGATTTTATTAACGTAT 1581
 QY 1002 TATTAAATTAAGCTTTT 1017
 Db 1582 TTTTGTGTAATTTT 1597

RESULT 14
 US-08-913-942-14
 ; Sequence 14, Application US/08913942
 ; Patent No. 6200578
 ; GENERAL INFORMATION:
 ; APPLICANT: St. Gene, Joseph
 ; APPLICANT: Barenkamp, Stephen J.
 ; TITLE OF INVENTION: HAEMOPHILUS ADHESION PROTEINS
 ; NUMBER OF SEQUENCES: 19
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Flehr Hohbach Test Albritton & Herbert LLP

; STREET: Four Embarcadero Center, Suite 3400
 ; CITY: San Francisco
 ; STATE: California
 ; COUNTRY: United States
 ; ZIP: 94111-4187
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA: US/08/913,942
 ; APPLICATION NUMBER: US/08/913,942
 ; FILING DATE: 29-DEC-1997
 ; CLASSIFICATION: 514
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/409,995
 ; FILING DATE: 24-MAR-1995
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: PCT/US96/4031
 ; FILING DATE: 22-MAR-1996
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Vance, Dolly A.
 ; REGISTRATION NUMBER: 39,054
 ; REFERENCE/DOCKET NUMBER: A-61053-1/RFT/RMS/DAV
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (415) 781-1989
 ; TELEFAX: (415) 398-3249
 ; TELEX: 910 277299
 ; INFORMATION FOR SEQ ID NO: 14:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 2037 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: unknown
 ; TOPOLOGY: unknown
 ; MOLECULE TYPE: DNA
 US-08-913-942-14

Query Match 3.7%; Score 43.4; DB 3; Length 2037;
 Best Local Similarity 55.7%; Pred. No. 0.31;
 Matches 83; Conservative 0; Mismatches 66; Indels 0; Gaps 0;
 QY 623 AATATCAATAAGTTGTCTTAAGAATTCGAAACTGGTTTGGCCAAACCAATCAGCATTTGTCT 682
 Db 1789 AAGTGAATGAATGTCGCAACGTCGAGATGTCAGTACTGCAAGTGCATTAGCGGCTTCA 1848
 QY 683 ATGTTAGTGCACCAATGGTGTAGGCAAAACGAGCGTTTCTGCTGCGGTAGGAGTTAT 742
 Db 1849 CAGTTACCAAGCCACTATGCCAGGTAATCAATGGTTTCTATTGCGGGAAGTAGTTAT 1908
 QY 743 AGAGATAAACTGATTCAGTCCATTGGTGT 771
 Db 1909 CAAGTCAAATGGTTAGCTATCGGGT 1937

RESULT 15
 US-09-268-347-25
 ; Sequence 25, Application US/09268347
 ; Patent No. 6335182
 ; GENERAL INFORMATION:
 ; APPLICANT: Loosmore, Sheena M.
 ; TITLE OF INVENTION: RECOMBINANT HAEMOPHILUS INFLUENZAE ADHESIN PROTEINS
 ; FILE REFERENCE: 1038-860
 ; CURRENT APPLICATION NUMBER: US/09/268,347
 ; CURRENT FILING DATE: 1999-03-16
 ; NUMBER OF SEQ ID NOS: 54
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 25
 ; LENGTH: 2079
 ; TYPE: DNA
 ; ORGANISM: Haemophilus influenzae
 US-09-268-347-25
 Query Match 3.7%; Score 43.4; DB 4; Length 2079;

Best Local Similarity 55.78; Pred. No. 0.31;
 Matches 83; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

Qy	623	AATATCAATAAGTTGTCTAAGAATTCGCAAACTGGTTTAGCCAACTCAGCATTGTCT	682
Db	1820	AAAGTGAATAAAGTGGCGCAACGTCAGATGCAGTACTGCAAGTGCATTAGCGGCTTCA	1879
Qy	683	ATGTTAGTGCACCAATGTTAGTGGCAAAACGAGCGTTTCTGCTGGGTAGGAGGTTAT	742
Db	1880	CAGTACCACAGCCACTATGCCAGGTAATCAATGGTTTCTATTGCGGAAGTAGTTAT	1939
Qy	743	AGAGATAAACTGCATTAGCCATTGGTGT	771
Db	1940	CAAGTCAAAATGTTTAGCTATCGGGT	1968

Search completed: May 15, 2004, 05:23:09
 Job time : 128 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 15, 2004, 01:31:40 ; Search time 561 Seconds
(without alignments)
8844.733 Million cell updates/sec

Title: US-10-030-529a-1
Perfect score: 1168
Sequence: 1 ataaacacgtcattgacatt.....aagccgttaagtgcgac 1168

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N Geneseq_29Jan04:*
1: geneseqn1980s:*
2: geneseqn1980s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2002s:*
6: geneseqn2001bs:*
7: geneseqn2003as:*
8: geneseqn2003bs:*
9: geneseqn2003cs:*
10: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1168	100.0	1168	4	Aaf25262 Nucleotid
2	1151	98.5	1189	4	Aaf25270 Nucleotid
3	1139	97.5	1197	4	Aaf25266 Nucleotid
4	1092.6	93.5	1205	4	Aaf25263 Nucleotid
5	1090	93.3	1231	4	Aaf25268 Nucleotid
6	866.8	74.2	1047	4	Aaf25269 Nucleotid
7	839.6	71.9	952	4	Aaf25264 Nucleotid
8	829.2	71.0	899	4	Aaf25265 Nucleotid
9	818	70.0	923	4	Aaf25267 Nucleotid
10	67.2	5.8	4426	3	Aac79587 Virulence
11	67.2	5.8	4426	6	Abq83464 Pasteurel
12	59.8	5.1	1095	3	Aaz54498 Neisseria
13	59.8	5.1	16526	3	Aaa81472 N. mening
14	59.8	5.1	110000	3	Continuation (7 of
15	59.8	5.1	172325	3	Aaf21613 Neisseria
16	59.8	5.1	349980	3	Aaf21612 Neisseria
17	59.2	5.1	972	7	Abx93390 Neisseria
18	57.6	4.9	1938	4	Aad17051 N. mening
19	57.6	4.9	1938	5	Aas43898 Neisseria
20	57.6	4.9	1947	4	Aad17043 Neisseria
21	57.6	4.9	1947	5	Aas43878 Neisseria
22	57.6	4.9	2388	4	Aad17035 N. mening
23	57.6	4.9	2388	5	Aas43870 Neisseria

24	57.6	4.9	2421	4	AAD17050	Aad17050 N. mening
25	57.6	4.9	2421	4	AAD17048	Aad17048 N. mening
26	57.6	4.9	2421	5	AAS43897	Aas43897 Neisseria
27	57.6	4.9	2421	5	AAS43895	Aas43895 Neisseria
28	57.6	4.9	2583	4	AAD17038	Aad17038 N. mening
29	57.6	4.9	2583	5	AAS43873	Aas43873 Neisseria
30	57.6	4.9	3798	4	AAD17041	Aad17041 N. mening
31	57.6	4.9	4335	4	AAD17052	Aad17052 N. mening
32	57.6	4.9	4335	5	AAS43899	Aas43899 Neisseria
33	57.6	4.9	4344	5	AAS43876	Aas43876 Neisseria
34	57.2	4.9	8139	7	ACA43016	Aca43016 Prokaryot
35	54	4.6	2000	7	ADA71938	Aad71938 Rice gene
36	54	4.6	110000	2	AAT42063_01	Continuation (2 of
37	53.2	4.6	2596	2	AAV41342	Aav41342 M. catarr
38	53.2	4.6	2673	2	AAV41346	Aav41346 M. catarr
39	53.2	4.6	4228	2	AAV41348	Aav41348 M. catarr
40	53.2	4.6	94750	4	AAF28551	Aaf28551 Genomic f
41	52.4	4.5	3295	2	AAV41344	Aav41344 M. catarr
42	52.2	4.5	558	7	AAD50226	Aad50226 Haemophil
43	51.8	4.4	18624	6	ABL33702	Ab133702 Human imm
44	50.8	4.3	9117	6	ABL33068	Ab133068 Human imm
45	49	4.2	18683	6	ABL32313	Ab132313 Human imm

ALIGNMENTS

RESULT 1
AAF25262
ID AAF25262 standard; DNA; 1168 BP.
XX
AC AAF25262;
XX
DT 30-APR-2001 (first entry)
XX
DE Nucleotide sequence of the DsrA locus.
XX
KW DsrA protein; outer membrane protein; serum resistance; vaccine;
KW chancroid disease; genital ulcer disease; ss.
XX
OS Haemophilus ducreyi.
PH Key Location/Qualifiers
FT CDS 101..874
FT /*tag= a
FT /product= "DsrA protein"
XX
PN WO200104138-A1.
XX
PD 18-JAN-2001.
XX
PF 07-JUL-2000; 2000WO-US018834.
XX
PR 09-JUL-1999; 99US-0143257P.
XX
PA (UYNC-) UNIV NORTH CAROLINA.
PA (ELKI/) ELKINS C.
XX
PI Elkins C;
XX
DR WPI; 2001-138311/14.
DR P-PSDB; AAB31705.
XX
PT Novel purified DsrA protein of Haemophilus ducreyi which confers serum
PT resistance to the bacteria used to produce vaccines that induce immune
PT response against the bacteria in subject at risk of developing chancroid.
XX
PS Claim 1; Fig 3; 80pp; English.
XX
CC The present sequence encodes a DsrA 30 kDa protein of Haemophilus
CC ducreyi. DsrA is an outer membrane protein that confers serum resistance
CC to the bacteria. DsrA antisense oligonucleotides are useful for detecting
CC a polynucleotide which encodes DsrA in a biological sample. The DsrA

CC polypeptide is used to produce vaccine compositions, which are useful for
 CC inducing a protective immune response in a subject at risk of developing
 CC Haemophilus ducreyi infection e.g. chancroid, a genital ulcer disease
 CC transmitted by sexual contact. DsrA, its catalytic or immunogenic
 CC fragments or peptides can be used for screening libraries of compounds in
 CC a variety of drug screening techniques. The proteins and peptides may be
 CC used as antigens in immunoassays for detection of Haemophilus ducreyi in
 CC various tissues and body fluids, e.g. blood, spinal fluid, sputum etc. .
 CC The nucleic acids are useful for the preparation of DsrA proteins
 XX

Seq Sequence 1168 BP; 357 A; 190 C; 209 G; 412 T; 0 U; 0 Other;

Query Match 100.0%; Score 1168; DB 4; Length 1168;
 Best Local Similarity 100.0%; Pred. No. 1.1e-243;
 Matches 1168; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATAAATAGCTGATTCACATTTTATGTAAGGTAGATAAGAAAGTAAATTCATATT 60
 DB 1 ATAAATAGCTGATTCACATTTTATGTAAGGTAGATAAGAAAGTAAATTCATATT 60
 QY 61 TACAATCAAGATTGACAAATTTTACTTAATGAGGTGATTATGAAATTTAAATGTTAGT 120
 DB 61 TACAATCAAGATTGACAAATTTTACTTAATGAGGTGATTATGAAATTTAAATGTTAGT 120
 QY 121 TGCCGTAGTGGGATTAGCTTCTTACTATTACAAATGGCTCAGCAGCGCCCAAGTT 180
 DB 121 TGCCGTAGTGGGATTAGCTTCTTACTATTACAAATGGCTCAGCAGCGCCCAAGTT 180
 QY 181 TGCTGGAGTATCTTTTGTATAGTATGATGATGATGATGATGATGATGATGATGATG 240
 DB 181 TGCTGGAGTATCTTTTGTATAGTATGATGATGATGATGATGATGATGATGATGATG 240
 QY 241 GTCTAAATGAAGCGGTTTCGATATTAAAGTCCAGGGATTAAATGAAGCCAAAGAAATG 300
 DB 241 GTCTAAATGAAGCGGTTTCGATATTAAAGTCCAGGGATTAAATGAAGCCAAAGAAATG 300
 QY 301 GATTTCTAAACAGGCTACTTATCTTGAATTACAGCATTATATGCTTATATCTCTGTCT 360
 DB 301 GATTTCTAAACAGGCTACTTATCTTGAATTACAGCATTATATGCTTATATCTCTGTCT 360
 QY 361 CGTGACATATGCTCTGCGGTTTCTCTAGCCCTATATGTTATATGCTGATGATCC 420
 DB 361 CGTGACATATGCTCTGCGGTTTCTCTAGCCCTATATGTTATATGCTGATGATCC 420
 QY 421 TGATCACTTGGATAAATCGCAGCAGCTGAAATGAAATTTGATGATTTTAAACGA 480
 DB 421 TGATCACTTGGATAAATCGCAGCAGCTGAAATGAAATTTGATGATTTTAAACGA 480
 QY 481 TTTAAGCAGCAGATTTTAAATTTAAAGTTCTTGTATGATGATGATTTTAAACAAA 540
 DB 481 TTTAAGCAGCAGATTTTAAATTTAAAGTTCTTGTATGATGATGATTTTAAACAAA 540
 QY 541 TATTGATATATAGTAAATTTTACTAGAACTGGGTACTTATTAGATGATTTCTTATCG 600
 DB 541 TATTGATATATAGTAAATTTTACTAGAACTGGGTACTTATTAGATGATTTCTTATCG 600
 QY 601 TATGATGCAACAAATACACATAATATCAATAAGTTGCTAAAGAAATTTGCAAACTGGTTT 660
 DB 601 TATGATGCAACAAATACACATAATATCAATAAGTTGCTAAAGAAATTTGCAAACTGGTTT 660
 QY 661 AGCCAAACCAATCAGATTTGCTATGTTAGTGCAACCAATGGTGTAGGCAAAACGAGCGT 720
 DB 661 AGCCAAACCAATCAGATTTGCTATGTTAGTGCAACCAATGGTGTAGGCAAAACGAGCGT 720
 QY 721 TTCTGCTGCGGTAGAGGTTATAGAGATAAAATGCAATTTAGGTTGTCGCTCAGC 780
 DB 721 TTCTGCTGCGGTAGAGGTTATAGAGATAAAATGCAATTTAGGTTGTCGCTCAGC 780
 QY 781 CATTACTGATCGTTTACCGCTTAAAGCGGTTAGCGTTCAATACCTACATGCGGCAT 840
 DB 781 CATTACTGATCGTTTACCGCTTAAAGCGGTTAGCGTTCAATACCTACATGCGGCAT 840
 QY 841 GTCTTATGGTCTTCTGTGTTATGAAATCTTAATCATTAGCTTTTAACTCACTAATCGTTT 900

DB 841 GTCTTATGGTCTTCTGTGTTATGAAATTTCTAATCATACGTTTAATCACTAATCGTTT 900
 QY 901 TGGTTATAATAAAGCGCTAAATGTTTCTCTCCTCACATTTAGCCCTTCTTATTTATCTTTG 960
 DB 901 TGGTTATAATAAAGCGCTAAATGTTTCTCTCCTCACATTTAGCCCTTCTTATTTATCTTTG 960
 QY 961 TTATAGCTTTTGTCTGTTTATAAACCCTGTTTATAGCCACTTTTATTAATTAAGCTTTTAAAG 1020
 DB 961 TTATAGCTTTTGTCTGTTTATAAACCCTGTTTATAGCCACTTTTATTAATTAAGCTTTTAAAG 1020
 QY 1021 CCTATTCAATCAGTTCTACTTTTCACTTTTTCACCATATTATTCGCCACTTCTAAAAACGG 1080
 DB 1021 CCTATTCAATCAGTTCTACTTTTCACTTTTTCACCATATTATTCGCCACTTCTAAAAACGG 1080
 QY 1081 TAATATTAACTGTTGTTAGCTTAAATGGGTACCTTCTATCGGAATTTTCTTAATGTTT 1140
 DB 1081 TAATATTAACTGTTGTTAGCTTAAATGGGTACCTTCTATCGGAATTTTCTTAATGTTT 1140
 QY 1141 CTAAATTAAGCCGTTTAAAGGTGCGGAC 1168
 DB 1141 CTAAATTAAGCCGTTTAAAGGTGCGGAC 1168

RESULT 2

AAF25270
 ID AAF25270 standard; DNA; 1189 BP.
 AC AAF25270;
 XX
 DT 30-APR-2001 (first entry)
 XX
 DE Nucleotide sequence of the DsrA protein from strain OF406.
 XX
 KW DsrA protein; outer membrane protein; serum resistance; vaccine;
 KW chancroid disease; genital ulcer disease; ss.
 XX
 OS Haemophilus ducreyi.
 XX
 PH Key Location/Qualifiers
 CDS 139..912
 FT /*tag= a
 FT /product= "DsrA protein"
 XX

W0200104138-A1.
 XX
 XX 18-JAN-2001.
 PD
 XX 07-JUL-2000; 2000WO-US018834.
 PF
 XX 09-JUL-1999; 99US-0143257P.
 PR
 XX (UNCL-) UNIV NORTH CAROLINA.
 PA (ELKI/) ELKINS C.
 XX
 PI Elkins C;
 XX
 XX WPI; 2001-138311/14.
 DR P-PSDB; AAB31713.
 DR

XX Novel purified DsrA protein of Haemophilus ducreyi which confers serum
 PT resistance to the bacteria used to produce vaccines that induce immune
 PT response against the bacteria in subject at risk of developing chancroid.
 XX

PS Claim 1; Page 58; 80pp; English.

XX The present sequence encodes a DsrA 30 kDa protein of Haemophilus
 CC ducreyi. DsrA is an outer membrane protein that confers serum resistance
 CC to the bacteria. DsrA antisense oligonucleotides are useful for detecting
 CC a polynucleotide which encodes DsrA in a biological sample. The DsrA
 CC polypeptide is used to produce vaccine compositions, which are useful for
 CC inducing a protective immune response in a subject at risk of developing
 CC Haemophilus ducreyi infection e.g. chancroid, a genital ulcer disease

transmitted by sexual contact. DsrA, its catalytic or immunogenic fragments or peptides can be used for screening libraries of compounds in a variety of drug screening techniques. The proteins and peptides may be used as antigens in immunoassays for detection of *Haemophilus ducreyi* in various tissues and body fluids, e.g. blood, spinal fluid, sputum etc. The nucleic acids are useful for the preparation of DsrA proteins

XX
SQ Sequence 1189 BP; 366 A; 188 C; 203 G; 432 T; 0 U; 0 Other;

Query Match 98.5%; Score 1151; DB 4; Length 1189;
Best Local Similarity 100.0%; Pred. No. 5.2e-240;
Matches 1151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATAAATACGTCATTGACATTTTATGTAAGTAGAATAAGAAAGTAAATCTATATT 60
DB |||||||
QY 39 ATAAATACGTCATTGACATTTTATGTAAGTAGAATAAGAAAGTAAATCTATATT 98
DB |||||||
QY 61 TACAATCAAGATTGACAAATTTATTTACTTAATGAGTGATTTGAAATTTAAATGTTT 120
DB |||||||
QY 99 TACAATCAAGATTGACAAATTTATTTACTTAATGAGTGATTTGAAATTTAAATGTTT 158
DB |||||||
QY 121 TGGCGTAGTGGGATTAGCTTGTCTACTATTACAAATCGCTCAGACGCGCAAGTT 180
DB |||||||
QY 159 TGGCGTAGTGGGATTAGCTTGTCTACTATTACAAATCGCTCAGACGCGCAAGTT 218
DB |||||||
QY 181 TGGCTGGAGTATCTTCTTTGATAGCTATGATGACTATGTAAGGTAAGTGGACTTG 240
DB |||||||
QY 219 TGGCTGGAGTATCTTCTTTGATAGCTATGATGACTATGTAAGGTAAGTGGACTTG 278
DB |||||||
QY 241 GTCTAATGAAGGGGGTTTCGATATTAAGTCCAGGGATTAATGAAGCAAAAGATG 300
DB |||||||
QY 279 GTCTAATGAAGGGGGTTTCGATATTAAGTCCAGGGATTAATGAAGCAAAAGATG 338
DB |||||||
QY 301 GATTCTAAACAGGCTACTTATCTTGAATTAAGCAATTAATGCTTTATCTCTGTCT 360
DB |||||||
QY 339 GATTCTAAACAGGCTACTTATCTTGAATTAAGCAATTAATGCTTTATCTCTGTCT 398
DB |||||||
QY 361 CGTGACATATGCTCTGGCGTTTCTCTAGCCCTATCTTATATCCGATGCTGATCC 420
DB |||||||
QY 399 CGTGACATATGCTCTGGCGTTTCTCTAGCCCTATCTTATATCCGATGCTGATCC 458
DB |||||||
QY 421 TGATCAACTTGGATAAATCGGACGAGCTGAAATTTGATAGTTTATTTTAAACGA 480
DB |||||||
QY 459 TGATCAACTTGGATAAATCGGACGAGCTGAAATTTGATAGTTTATTTAAACGA 518
DB |||||||
QY 481 TTTAAGACACGATTTTAAATTAAGTTCTTGATGACGATTTTCCAAAATAAACAATA 540
DB |||||||
QY 519 TTTAAGACACGATTTTAAATTAAGTTCTTGATGACGATTTTCCAAAATAAACAATA 578
DB |||||||
QY 541 TATTGATCTATAAGTAAATTTTACTAGAACTGGGTACTTATTAGATGATCTTATCG 600
DB |||||||
QY 579 TATTGATCTATAAGTAAATTTTACTAGAACTGGGTACTTATTAGATGATCTTATCG 638
DB |||||||
QY 601 TATTGATGGAACAAATACACATATATCAATTAAGTTCTTAAAGAAATTCGAAATCG 660
DB |||||||
QY 639 TATTGATGGAACAAATACACATATATCAATTAAGTTCTTAAAGAAATTCGAAATCG 698
DB |||||||
QY 661 AGCCAAACCAATCAGATTTGCTATGTTAGTGAACCAAAATGGTGTAGGCAAAACGCGT 720
DB |||||||
QY 699 AGCCAAACCAATCAGATTTGCTATGTTAGTGAACCAAAATGGTGTAGGCAAAACGCGT 758
DB |||||||
QY 721 TTCTGCTGCGGTAGGAGTTATAGAGATAAATCTGCATTTAGCCATTTGGTGTGCGCTC 780
DB |||||||
QY 759 TTCTGCTGCGGTAGGAGTTATAGAGATAAATCTGCATTTAGCCATTTGGTGTGCGCTC 818
DB |||||||
QY 781 CATTTACTGATCGTTTACCGCTAAAGCGGTGTAGCGTTCAATACCTACATGCGGCAT 840
DB |||||||
QY 819 CATTTACTGATCGTTTACCGCTAAAGCGGTGTAGCGTTCAATACCTACATGCGGCAT 878
DB |||||||
QY 841 GTCTTATGCTGCTTCTGTTGTTATGAATTTCAATCAATTAAGTTTAACTAATCGTTT 900
DB |||||||
QY 879 GTCTTATGCTGCTTCTGTTGTTATGAATTTCAATCAATTAAGTTTAACTAATCGTTT 938
DB |||||||

QY 901 TGGTTATAATAAAAGCGCTAAATGTTTCTCCTCACATTTAGCCTTTCTTATTTATCTTTG 960
DB |||||||
QY 939 TGGTTATAATAAAAGCGCTAAATGTTTCTCCTCACATTTAGCCTTTCTTATTTATCTTTG 998
DB |||||||
QY 961 TTATAGCTTTTGTCTGTATATAAAACCGTTTATTAGCCACTTTTATTAATTAAGCTTTTAAAG 1020
DB |||||||
QY 999 TTATAGCTTTTGTCTGTATATAAAACCGTTTATTAGCCACTTTTATTAATTAAGCTTTTAAAG 1058
DB |||||||
QY 1021 CCTATTCAATCAGTTCTCTACTTTTACCTTTTCCACCATATTTATCCGCCACTTCTTAAACCG 1080
DB |||||||
QY 1059 CCTATTCAATCAGTTCTCTACTTTTACCTTTTCCACCATATTTATCCGCCACTTCTTAAACCG 1118
DB |||||||
QY 1081 TAATATAAGTTGGTTTAGCTTAATTTGGTACCTTCTATCGGAATTTTCTTAAATGTT 1140
DB |||||||
QY 1119 TAATATAAGTTGGTTTAGCTTAATTTGGTACCTTCTATCGGAATTTTCTTAAATGTT 1178
DB |||||||
QY 1141 CTAAATTAAG 1151
DB |||||||
QY 1179 CTAAATTAAG 1189
DB |||||||

RESULT 3
AAF25266
ID AAF25266 standard; DNA; 1197 BP.
XX
AC AAF25266;
XX
DT 30-APR-2001 (first entry)
XX
DE Nucleotide sequence of the DsrA protein from strain CIP542 (CDC).
XX
KW DsrA protein; outer membrane protein; serum resistance; vaccine;
KW chancroid disease; genital ulcer disease; ss.
XX
OS *Haemophilus ducreyi*.
XX
FH Key Location/Qualifiers
FT CDS 147..920
FT /tag= a
FT /product= "DsrA protein"
XX
WO200104138-A1.
XX
PD 18-JAN-2001.
XX
PF 07-JUL-2000; 2000WO-US018834.
XX
PR 09-JUL-1999; 99US-0143257P.
XX
PA (UNIC-) UNIV NORTH CAROLINA.
PA (ELKI/) ELKINS C.
XX
PI Elkins C;
XX
WPI; 2001-138311/14.
DR P-PSDB; AAB31709.
XX
PT Novel purified DsrA protein of *Haemophilus ducreyi* which confers serum resistance to the bacteria used to produce vaccines that induce immune response against the bacteria in subject at risk of developing chancroid.
XX
PS Claim 1; Page 55; 80pp; English.
XX
CC The present sequence encodes a DsrA 30 kDa protein of *Haemophilus ducreyi*. DsrA is an outer membrane protein that confers serum resistance to the bacteria. DsrA antisense oligonucleotides are useful for detecting a polynucleotide which encodes DsrA in a biological sample. The DsrA polypeptide is used to produce vaccine compositions, which are useful for inducing a protective immune response in a subject at risk of developing *Haemophilus ducreyi* infection e.g. chancroid, a genital ulcer disease transmitted by sexual contact. DsrA, its catalytic or immunogenic fragments or peptides can be used for screening libraries of compounds in a variety of drug screening techniques. The proteins and peptides may be

CC used as antigens in immunoassays for detection of Haemophilus ducreyi in
CC various tissues and body fluids, e.g. blood, spinal fluid, sputum etc. .
CC The nucleic acids are useful for the preparation of DsrA proteins
XX
SQ Sequence 1197 BP; 369 A; 191 C; 204 G; 433 T; 0 U; 0 Other;
Query Match 97.5%; Score 1139; DB 4; Length 1197;
Best Local Similarity 99.9%; Pred. No. 2.1e-237;
Matches 1150; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
Qy 1 ATAAATACGTCATTCAGACATTTTAAATGTAAGTAGAATAAGAAAGTAATCTATATT 60
Db 47 ATAAATACGTCATTCAGACATTTTAAATGTAAGTAGAATAAGAAAGTAATCTATATT 105
Qy 61 TACAATCAAGATTGACAAATTTTACCTTAATGAGGTGATTAATAAATTAATGTTAGT 120
Db 107 TACAATCAAGATTGACAAATTTTACCTTAATGAGGTGATTAATAAATTAATGTTAGT 166
Qy 121 TGCGTAGTGGGATTAGCTTGTCTTACTATTACAACAATGGCTCAGCAGCGCCCAAGTT 180
Db 167 TGCGTAGTGGGATTAGCTTGTCTTACTATTACAACAATGGCTCAGCAGCGCCCAAGTT 225
Qy 181 TGCTGGAGTATCTCTTTGTATAGCTATGATGATGACTATGGTAAGGGTAAATGGACTTG 240
Db 227 TGCTGGAGTATCTCTTTGTATAGCTATGATGATGACTATGGTAAGGGTAAATGGACTTG 285
Qy 241 GTCTAATGAAGCGGTTTCGATATTAAAGTCCAGGGATTAAATGAAGCCAAAGAAATG 300
Db 287 GTCTAATGAAGCGGTTTCGATATTAAAGTCCAGGGATTAAATGAAGCCAAAGAAATG 346
Qy 301 GATTTCTAAACAGGCTACTTATCTTGAATTACAGCATTATATGCTTATATCTCTGTTCT 360
Db 347 GATTTCTAAACAGGCTACTTATCTTGAATTACAGCATTATATGCTTATATCTCTGTTCT 406
Qy 361 CGTGACATATGCTCCTGGCGTTTCTCCTAGCCCTATAGTATATCCGATGTCGTATCC 420
Db 407 CGTGACATATGCTCCTGGCGTTTCTCCTAGCCCTATAGTATATCCGATGTCGTATCC 465
Qy 421 TGATCAACTTGGATAAATCGGCAGCAGCTGAAATGAAATTTGTATAGTATTATTTAACA 480
Db 467 TGATCAACTTGGATAAATCGGCAGCAGCTGAAATTTGTATAGTATTATTTAACA 526
Qy 481 TTTAAGCACCATTTTAAATTTAAAGTTCTTGTAGTCACGCTATTTCCAAAATAAACA 540
Db 527 TTTAAGCACCATTTTAAATTTAAAGTTCTTGTAGTCACGCTATTTCCAAAATAAACA 586
Qy 541 TATTGATACATAAGTAATTTTACTAGAACTGGGTACTTATTTAGATGATCTTATCG 600
Db 587 TATTGATACATAAGTAATTTTACTAGAACTGGGTACTTATTTAGATGATCTTATCG 646
Qy 601 TATGATGAACAAATACACATAATATCAATAAGTTGTCTAAAGAAATTCGAACTGGTTT 660
Db 647 TATGATGAACAAATACACATAATATCAATAAGTTGTCTAAAGAAATTCGAACTGGTTT 706
Qy 661 AGCCAACCAATCAGCATTTGTATGTTAGTGCACAAATGGTGTAGGCCAAACGAGCGT 720
Db 707 AGCCAACCAATCAGCATTTGTATGTTAGTGCACAAATGGTGTAGGCCAAACGAGCGT 765
Qy 721 TTTCTGCTGCGTAGAGGTTATAGAGATAAAATGCAATAGCCATTTGGTGTGCGCTCAAG 780
Db 767 TTTCTGCTGCGTAGAGGTTATAGAGATAAAATGCAATAGCCATTTGGTGTGCGCTCAAG 826
Qy 781 CATTACTGATCGCTTACCGCTAAAGCGGTGAGCGTTCAATACCTACATGCGGCAT 840
Db 827 CATTACTGATCGCTTACCGCTAAAGCGGTGAGCGTTCAATACCTACATGCGGCAT 886
Qy 841 GTCTTATGGTCTCTCTGTTGTTATGAATTTCTAATCATTAGCTTTAATCACTAATCGTTT 900
Db 887 GTCTTATGGTCTCTCTGTTGTTATGAATTTCTAATCATTAGCTTTAATCACTAATCGTTT 946
Qy 901 TGGTTATATAAAGGCTAAATGTTTCTCCTCAGATTTAGCCCTTCTTATTTATCTTTG 960
Db 947 TGGTTATATAAAGGCTAAATGTTTCTCCTCAGATTTAGCCCTTCTTATTTATCTTTG 1006

Qy 961 TTATAG-CTTTTCTGCTGTATATAAACCGTTTTTTTAGCCACCTTTTATTAATTAAGCTTTTAA 1019
Db 1007 TTATAGCTTTTCTGCTGTATATAAACCGTTTTTTTAGCCACCTTTTATTAATTAAGCTTTTAA 1066
Qy 1020 GCCTATTCAATCAGTCTTACTTTTCACCTTTTTCACCATATATCCGCCACTTCTAAACG 1079
Db 1067 GCCTATTCAATCAGTCTTACTTTTCACCTTTTTCACCATATATCCGCCACTTCTAAACG 1126
Qy 1080 GTAATATTAGTGGTTTTCAGCCTAAATTTGGGTACCTTCTATCGAAATTTTCTTAAATGT 1139
Db 1127 GTAATATTAGTGGTTTTCAGCCTAAATTTGGGTACCTTCTATCGAAATTTTCTTAAATGT 1186
Qy 1140 TCTAAATAATTA 1150
Db 1187 TCTAAATAATTA 1197
RESULT 4
AAF25263
ID AAF25263 standard; DNA; 1205 BP.
XX AAF25263;
DT 30-APR-2001 (first entry)
XX
DE Nucleotide sequence of the DsrA protein from strain CIPA75.
XX
KW DsrA protein; outer membrane protein; serum resistance; vaccine;
KM chancroid disease; genital ulcer disease; ss.
XX
OS Haemophilus ducreyi.
XX
FH Key Location/Qualifiers
FT CDS 135..929
FT /*tag= a
FT /product= "DsrA protein"
XX
PN W0200104138-A1.
XX
PD 18-JAN-2001.
XX
PF 07-JUL-2000; 2000WO-US018834.
XX
PR 09-JUL-1999; 99US-0143257P.
XX
PA (UNCL-) UNIV NORTH CAROLINA.
PA (ELKI/) ELKINS C.
XX
PI Elkins C;
XX
XX WPI; 2001-138311/14.
XX
PT Novel purified DsrA protein of Haemophilus ducreyi which confers serum
PT resistance to the bacteria used to produce vaccines that induce immune
PT response against the bacteria in subject at risk of developing chancroid.
XX
PS Claim 1; Page 53; 80pp; English.
XX
XX The present sequence encodes a DsrA 30 kDa protein of Haemophilus
CC ducreyi. DsrA is an outer membrane protein that confers serum resistance
CC to the bacteria. DsrA antisense oligonucleotides are useful for detecting
CC a polynucleotide which encodes DsrA in a biological sample. The DsrA
CC polypeptide is used to produce vaccine compositions, which are useful for
CC inducing a protective immune response in a subject at risk of developing
CC Haemophilus ducreyi infection e.g. chancroid, a genital ulcer disease
CC transmitted by sexual contact. DsrA, its catalytic or immunogenic
CC fragments or peptides can be used for screening libraries of compounds in
CC a variety of drug screening techniques. The proteins and peptides may be
CC used as antigens in immunoassays for detection of Haemophilus ducreyi in
CC various tissues and body fluids, e.g. blood, spinal fluid, sputum etc. .
XX The nucleic acids are useful for the preparation of DsrA proteins

```
SQ Sequence 1205 BP; 380 A; 190 C; 200 G; 435 T; 0 U; 0 Other;
Query Match 93.5%; Score 1092.6; DB 4; Length 1205;
Best Local Similarity 97.4%; Pred. No. 2.4e-227;
Matches 1141; Conservative 0; Mismatches 4; Indels 26; Gaps 2;

QY 1 ATAATAACGTCATTGACATTTTAAATGTAAGGTAGAAATAAGAAAGTAAATTCATATT 60
DB 40 ATAATAACGTCATTGACATTTTAAATGTAAGGTAGAAATAAGAAAGTAAATTCATATT 94
QY 61 TACAATCAAGATTGACAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 120
DB 95 TACAATCAAGATTGACAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 154
QY 121 TCGCGTAGTGGATTAGCTTCTTCTACTATTACCAATGGCTCAGCAGCGCCGCAAGTT 180
DB 155 TCGCGTAGTGGATTAGCTTCTTCTACTATTACCAATGGCTCAGCAGCGCCGCAAGTT 214
QY 181 TCGTGGAGTAGTCTTCTTCTTGTATAGCTATGAGTATGACTATGTAAGGTAAATGGACTTG 240
DB 215 TCGTGGAGTAGTCTTCTTCTTGTATAGCTATGAGTATGACTATGTAAGGTAAATGGACTTG 274
QY 241 GTCTAATGAAGCGGTTTCGATATTAAGTGCAGGATTAATAATGAAGCCAAAGAAATG 300
DB 275 GTCTAATGAAGCGGTTTCGATATTAAGTGCAGGATTAATAATGAAGCCAAAGAAATG 334
QY 301 GATTTCTAAACAGCTACTTATCTTGAATTAACACATTAATGCTTATACCTCTGTCT 360
DB 335 GATTTCTAAACAGCTACTTATCTTGAATTAACACATTAATGCTTATACCTCTGTCT 394
QY 361 CGTCACATATGCTCTCGCGGTTCTCTAGCCCTATCTATGTTATCGATGCTGATCC 420
DB 395 CGTCACATATGCTCTCGCGGTTCTCTAGCCCTATCTATGTTATCGATGCTGATCC 454
QY 421 TGATCAACTTGGAAATAATCGGCAGCAGCTGAAATTTGAATTTGATAGTTATTTAAACGA 480
DB 455 TGATCAACTTGGAAATAATCGGCAGCAGCTGAAATTTGAATTTGATAGTTATTTAAACGA 514
QY 481 TTTAAGACAGATTTTAAATTTAAAGTTCTGTAGTCAGGATTTTCCAAAATAAACAAAA 540
DB 515 TTTAAGACAGATTTTAAATTTAAAGTTCTGTAGTCAGGATTTTCCAAAATAAACAAAA 574
QY 541 TATTGATACATAAGTAAATTTTACTAGACCTGGTACTTATTTAGATGATTCCTATCG 600
DB 575 TATTGATACATAAGTAAATTTTACTAGACCTGGTACTTATTTAGATGATTCCTATCG 634
QY 601 TATGATGGAAC-----AAAATACACATAATATCAATAAATACACATAATCAATAAGTTGTC 639
DB 635 TATGATGGAACAAATACACATAATATCAATAAATACACATAATCAATAAGTTGTC 694
QY 640 TAAAGAAATGCAAACTGGTTTACCAACCAATCAGCATTTGTCTATGTAGTGCACCAAA 699
DB 695 TAAAGAAATGCAAACTGGTTTACCAACCAATCAGCATTTGTCTATGTAGTGCACCAAA 754
QY 700 TGGTGTAGGCAAAACGAGCGTTCTCTGCGGTAGGAGGTATAGATAAACCTGCATT 759
DB 755 TGGTGTAGGCAAAACGAGCGTTCTCTGCGGTAGGAGGTATAGATAAACCTGCATT 814
QY 760 AGCCATTGGTGTGGCTCAGCATTTACTGATGCTTTTACCGCTAAAGCGGTGAGCGTT 819
DB 815 AGCCATTGGTGTGGCTCAGCATTTACTGATGCTTTTACCGCTAAAGCGGTGAGCGTT 874
QY 820 CAATACCTACAAATCGCGCATGTCTTATGFGCTTCTGTTGGTTATGAATCTTAATCATT 879
DB 875 CAATACCTACAAATCGCGCATGTCTTATGFGCTTCTGTTGGTTATGAATCTTAATCATT 934
QY 880 ACGTTTAACTACTAATCGTTTGTGTTTAAATAAAGCTAAATGTTTCTCTCCTACATTT 939
DB 935 ACGTTTAACTACTAATCGTTTGTGTTTAAATAAAGCTAAATGTTTCTCTCCTACATTT 994
QY 940 ACGCTTTCTTATTTATCTTCTGTTATAGCTTTGCTGTTTAAACCGTTTTTTAGCCACT 999
DB 995 ACGCTTTCTTATTTATCTTCTGTTATAGCTTTGCTGTTTAAACCGTTTTTTAGCCACT 1054
```

```
QY 1000 TTTATTAAATTAAGCTTTTAAAGCCTATTCAATCAGTTCTACTTTTCACTTTTCCACCATAT 1059
DB 1055 TTTATTAAATTAAGCTTTTAAAGCCTATTCAATCAGTTCTACTTTTCACTTTTCCACCATAT 1114
QY 1060 TATCCCGCACCCTTCTAAACCGTAAATTAATTAAGTTGGTTTAGCCCTAAATTTGGGTACCTTTCTA 1119
DB 1115 TATCCCGCACCCTTCTAAACCGTAAATTAATTAAGTTGGTTTAGCCCTAAATTTGGGTACCTTTCTA 1174
QY 1120 TCGGAATTTTCTTAAATGTTCTAAATTTAA 1150
DB 1175 TCGGAATTTTCTTAAATGTTCTAAATTTAA 1205

RESULT 5
AAF25288
ID AAF25268 standard; DNA; 1231 BP.
XX
AC AAF25268;
XX
DT 30-APR-2001 (first entry)
XX
Nucleotide sequence of the DsrA protein from strain V-1157.
XX
DsrA protein; outer membrane protein; serum resistance; vaccine;
chancroid disease; genital ulcer disease; ss.
XX
Haemophilus ducreyi.
OS
FH Key Location/Qualifiers
FT CDS 140..955
FT /tag= a
FT /product= "DsrA protein"
XX
WO200104138-A1.
PN
XX
XX 18-JAN-2001.
XX
XX 07-JUL-2000; 2000WO-US018834.
XX
XX 09-JUL-1999; 99US-0143257P.
XX
(DVNC-) UNIV NORTH CAROLINA.
PA (ELKI/) ELKINS C.
XX
Elkins C;
XX
XX WPI; 2001-138311/14.
DR P-PSDB; AAB31711.
XX
XX
Novel purified DsrA protein of Haemophilus ducreyi which confers serum
resistance to the bacteria used to produce vaccines that induce immune
response against the bacteria in subject at risk of developing chancroid.
XX
Claim 1; Page 56-57; 80pp; English.
XX
The present sequence encodes a DsrA 30 kDa protein of Haemophilus
ducreyi. DsrA is an outer membrane protein that confers serum resistance
to the bacteria. DsrA antisense oligonucleotides are useful for detecting
a polynucleotide which encodes DsrA in a biological sample. The DsrA
polypeptide is used to produce vaccine compositions, which are useful for
inducing a protective immune response in a subject at risk of developing
Haemophilus ducreyi infection e.g. chancroid, a genital ulcer disease
transmitted by sexual contact. DsrA, its catalytic or immunogenic
fragments or peptides can be used for screening libraries of compounds in
a variety of drug screening techniques. The proteins and peptides may be
used as antigens in immunoassays for detection of Haemophilus ducreyi in
various tissues and body fluids, e.g. blood, spinal fluid, sputum etc.
XX
The nucleic acids are useful for the preparation of DsrA proteins
SQ Sequence 1231 BP; 391 A; 196 C; 202 G; 442 T; 0 U; 0 Other;
Query Match 93.3%; Score 1090; DB 4; Length 1231;
```

Best Local Similarity 96.1%; Pred. No. 8.8e-227;		Matches 1145; Conservative 0; Mismatches 5; Indels 42; Gaps 1;	
Qy	1	ATAAATACGTCATTTGACATATTTTAAATGTAAGGTAGAATAAGAAAGTAAATTTCTATATT	60
Db	40	ATAAATACGTCATTTGACATATTTTAAATGTAAGGTAGAATAAGAAAGTAAATTTCTATATT	99
Qy	61	TACAATCAAGATGACAAATATTTTACTTTAATGAGGTGATTATGAAAAATTAATATTTAGT	120
Db	100	TACAATCAAGATGACAAATATTTTACTTTAATGAGGTGATTATGAAAAATTAATATTTAGT	159
Qy	121	TGCCGTAGTGGATAGCTTGTCTTACTATATTAACAATGGCTCAGACGCCGCAAGTT	180
Db	160	TGCCGTAGTGGATAGCTTGTCTTACTATATTAACAATGGCTCAGACGCCGCAAGTT	219
Qy	181	TGCTGGAGTATCTTCTTTGTATAGCTATGAGTATGACTATGTAAGGTAATGACCTTG	240
Db	220	TGCTGGAGTATCTTCTTTGTATAGCTATGAGTATGACTATGTAAGGTAATGACCTTG	279
Qy	241	GTCTAATGAAGCGGTTTCGATATTAAGTCCAGGATTAATAAGCAAGCAAGATG	300
Db	280	GTCTAATGAAGCGGTTTCGATATTAAGTCCAGGATTAATAAGCAAGCAAGATG	339
Qy	301	GATTTCTAAACAGGCTACTTATCTTGAATACAGCATTAATGCTTATATCTCTGTTCT	360
Db	340	GATTTCTAAACAGGCTACTTATCTTGAATACAGCATTAATGCTTATATCTCTGTTCT	399
Qy	361	CGTGACATATCTCTCTGCGTTCTCTAGCCCTATCTGTTATATCGATGCTGATCC	420
Db	400	CGTGACATCTCTCTGAGCTCTCTAGCTCTATCTGTTATCTGATGCTGATCC	459
Qy	421	TGATCAACTTGGAAATAATCGGCAGCAGCTGAAATTTGATAGTTATTTTAAACGA	480
Db	460	TGATCAACTTGGAAATAATCGGCAGCAGCTGAAATTTGATAGTTATTTTAAACGA	519
Qy	481	TTTAAGACACATTTTAAATTTAAAGTCTTGATGACAGGATTTTCCAAAAATAACAAA	540
Db	520	TTTAAGACACATTTTAAATTTAAAGTCTTGATGACAGGATTTTCCAAAAATAACAAA	579
Qy	541	TATTGATCTATAAGTAAATTTTACTAGACTGGTACTTTTATAGATGCTTCTATCG	600
Db	580	TATTGATCTATAAGTAAATTTTACTAGAACTGGTACTTTTATAGATGCTTCTATCG	639
Qy	601	TATGATGGAAC-----AAAATAC	618
Db	640	TATGATGGAACAAATACACATATATCAATAAATAACATATATCAATAAATAAC	699
Qy	619	ACATAATATCAATAGTTGCTAAAGATTTGCAAACTGGTTTAGCCAAACCAATCAGATT	678
Db	700	ACATAATATCAATAGTTGCTAAAGATTTGCAAACTGGTTTAGCCAAACCAATCAGATT	759
Qy	679	GTCTATGTTAGTCAACCAATGTTAGTGGCAAAACGAGCGTTTCTGCGGTAGGAGG	738
Db	760	GTCTATGTTAGTCAACCAATGTTAGTGGCAAAACGAGCGTTTCTGCGGTAGGAGG	819
Qy	739	TTATAGAGATAAACTGCATTAGCCATTGGTGTGCGGCTCAGCATTTACTGATCGCTTAC	798
Db	820	TTATAGAGATAAACTGCATTAGCCATTGGTGTGCGGCTCAGCATTTACTGATCGCTTAC	879
Qy	799	CGCTAAAGCGGGTGTAGCGTTCAATACCTAATGCGGCGATGCTTATGCGTCTTGT	858
Db	880	CGCTAAAGCGGGTGTAGCGTTCAATACCTAATGCGGCGATGCTTATGCGTCTTGT	939
Qy	859	TGGTTATGAATCTTAATCATTCAGTTTAACTCACTAATCGTTTGGTTATATAAAGGC	918
Db	940	TGGTTATGAATCTTAATCATTCAGTTTAACTCACTAATCGTTTGGTTATATAAAGGC	999
Qy	919	TAAATGTTTCCTCCACATTTAGCCTTCTTATTTATCTTTGTTATAGCTTTTGTGTTA	978
Db	1000	TAAATGTTTCCTCCACATTTAGCCTTCTTATTTATCTTTGTTATAGCTTTTGTGTTA	1059
Qy	979	TAAACCGTTTTTTAGCCACTTTTATTAATTAAGCTTTTAAAGCTTATCAATCAGTTCTA	1038

Db	1060	TAAAACGGTTTTTTAGCCACTTTTATTAATAAGCTTTTAAAGCCTATTCAATCAGTTCTA	1119
Qy	1039	CTTTCACTTTTTTACCATATTTATCGCCACTTCTTAAACGGTAATATTAAAGTTGGTTTA	1098
Db	1120	CTTTCACTTTTTTACCATATTTATCGCCACTTCTTAAACGGTAATATTAAAGTTGGTTTA	1179
Qy	1099	GCTAAATGGGTACCTTCTATCGGAATTTTTTTCTAAATGTTCTAAAAATTAA	1150
Db	1180	GCTAAATGGGTACCTTCTATCGGAATTTTTTTCTAAATGTTCTAAAAATTAA	1231
RESULT 6			
AAF25269			
ID	AAF25269 standard; DNA; 1047 BP.		
XX			
AC	AAF25269;		
XX			
DT	30-APR-2001 (first entry)		
XX			
DE	Nucleotide sequence of the DsrA protein from strain M90-02.		
XX			
KW	DsrA protein; outer membrane protein; serum resistance; vaccine;		
KW	chancroid disease; genital ulcer disease; ss.		
XX	Haemophilus ducreyi.		
OS			
FH	Key	Location/Qualifiers	
FT	CDS	140..961	
FT		/*tag= a	
FT		/product= "DsrA protein"	
XX	WO200104138-A1.		
XX			
PD	18-JAN-2001.		
XX			
PF	07-JUL-2000; 2000WO-US018834.		
XX			
PR	09-JUL-1999; 99US-0143257P.		
XX			
PA	(UYN-) UNIV NORTH CAROLINA.		
PA	(ELKI/) ELKINS C.		
XX			
PI	Elkins C;		
XX			
DR	WPI; 2001-138311/14.		
DR	P-PSDB; AAB31712.		
XX			
PT	Novel purified DsrA protein of Haemophilus ducreyi which confers serum		
PT	resistance to the bacteria used to produce vaccines that induce immune		
PT	response against the bacteria in subject at risk of developing chancroid.		
XX			
PS	Claim 1; Page 57-58; 80pp; English.		
XX			
CC	The present sequence encodes a DsrA 30 kDa protein of Haemophilus		
CC	ducreyi. DsrA is an outer membrane protein that confers serum resistance		
CC	to the bacteria. DsrA antisense oligonucleotides are useful for detecting		
CC	a polynucleotide which encodes DsrA in a biological sample. The DsrA		
CC	polypeptide is used to produce vaccine compositions, which are useful for		
CC	inducing a protective immune response in a subject at risk of developing		
CC	Haemophilus ducreyi infection e.g. chancroid, a genital ulcer disease		
CC	transmitted by sexual contact. DsrA, its catalytic or immunogenic		
CC	fragments or peptides can be used for screening libraries of compounds in		
CC	a variety of drug screening techniques. The proteins and peptides may be		
CC	used as antigens in immunoassays for detection of Haemophilus ducreyi in		
CC	various tissues and body fluids, e.g. blood, spinal fluid, sputum etc. .		
CC	The nucleic acids are useful for the preparation of DsrA proteins		
XX			
SQ	Sequence 1047 BP; 339 A; 163 C; 180 G; 365 T; 0 U; 0 Other;		

Query Match 74.2%; Score 866.8; DB 4; Length 1047;
Best Local Similarity 94.4%; Pred. No. 2.2e-178;
Matches 950; Conservative 0; Mismatches 7; Indels 49; Gaps 3;


```
QY 3 AAATACGTCATTGACATTTTAAATGTAAGGTAGAATAAGAAAGTAAATTTCTATATTTA 62
DB |||
QY 42 AATACCGTCATTGACATTTTAAATGTAAGGTAGAATAAGAAAGTAAATTTCTATATTTA 101
DB |||
QY 63 CAATCAAGATGACAAATTTTAAATGTAAGGTAGAATAAGAAAGTAAATTTCTATATTTA 122
DB |||
QY 102 CAATCAAGATGACAAATTTTAAATGTAAGGTAGAATAAGAAAGTAAATTTCTATATTTA 161
DB |||
QY 123 CCGTAGTGGGATGAGTCTTCTACTATTCAACAATGGCTCAGCAGCGCCAAAGTTTG 182
DB |||
QY 162 CCGTAGTGGGATGAGTCTTCTACTATTCAACAATGGCTCAGCAGCGCCAAAGTTTG 221
DB |||
QY 183 CTGGAGTATCTTTCTGTATAGCTATGAGTATGACTATGTAAGGGTAAATTTCTATATTTA 242
DB |||
QY 222 CTGGAGTATCTTTCTGTATAGCTATGAGTATGACTATGTAAGGGTAAATTTCTATATTTA 281
DB |||
QY 243 CTAATGAAGCGGTTTCGATATTAAGTGCAGCGGATTAATAATGAAGCCAAAGATGGA 302
DB |||
QY 282 CTAATGAAGCGGTTTCGATATTAAGTGCAGCGGATTAATAATGAAGCCAAAGATGGA 341
DB |||
QY 303 TTTCTAACAGGCTACTTCTTGAATTAAGTGCAGCGGATTAATAATGAAGCCAAAGATGGA 362
DB |||
QY 342 TTTCTAACAGGCTACTTCTTGAATTAAGTGCAGCGGATTAATAATGAAGCCAAAGATGGA 401
DB |||
QY 363 TGACATATGCTCTCGGCTTCTCTCTAG-----CCCTATGCTGTATATCCGATGCTG 416
DB |||
QY 402 TGACATCTGCTCTCGGCTTCTCTCTAGCTCTATCTCTATCTGTATATCCGATGCTG 461
DB |||
QY 417 ATCTGTATCAACTTGGAAATAATCGGAGCAGCTGAAATTTGAATTTGTATAGTTTATTTA 476
DB |||
QY 462 ATCTGTATCAACTTGGAAATAATCGGAGCAGCTGAAATTTGAATTTGTATAGTTTATTTA 521
DB |||
QY 477 ACGATTTAAGACAGATTTAAATTAAGTCTTGTATGACGATTTTCCAAATAAATCAAC 536
DB |||
QY 522 ACGATTTAAGACAGATTTAAATTAAGTCTTGTATGACGATTTTCCAAATAAATCAAC 581
DB |||
QY 537 AAAATATTGTATCTATTAAGTAAATTTTCTAGAACTGGGTACTTATTTAGATGATCTT 596
DB |||
QY 582 AAAATATTGTATCTATTAAGTAAATTTTCTAGAACTGGGTACTTATTTAGATGATCTT 641
DB |||
QY 597 ATCGTATGATGGAAC-----AAA 614
DB |||
QY 642 ATCGTATGATGGAACAAATACACATAATATCAATAAAATACACATAATATCAATAAAA 701
DB |||
QY 615 ATACACATAATATCAATAGTTGTCTAAAGAAATTTGAAATCTGTTTATGACCAACCAATCAG 674
DB |||
QY 702 ATACACATAATATCAATAGTTGTCTAAAGAAATTTGAAATCTGTTTATGACCAACCAATCAG 761
DB |||
QY 675 CATTGCTATGTTAGTGCACCAATGTTGTAGGCAAAACGAGGTTTCTGCTGCGGTAG 734
DB |||
QY 762 CATTGCTATGTTAGTGCACCAATGTTGTAGGCAAAACGAGGTTTCTGCTGCGGTAG 821
DB |||
QY 735 GAGGTTATAGAGATAAAATGTCATTAGTCATTGTTGTCGGCTCAGCGCATTTACTGATCGCT 794
DB |||
QY 822 GAGGTTATAGAGATAAAATGTCATTAGTCATTGTTGTCGGCTCAGCGCATTTACTGATCGCT 881
DB |||
QY 795 TTACCGTTAAAGCGGTTGAGCGTTCAATACCTTCAATAGCGGCAATGCTTATGTTGCTT 854
DB |||
QY 882 TTACCGTTAAAGCGGTTGAGCGTTCAATACCTTCAATAGCGGCAATGCTTATGTTGCTT 941
DB |||
QY 855 CTGTTGTTATGAATTTCAATCATTTAGTTTAAATCACTATCTGTTTGGTTTATATAAAA 914
DB |||
QY 942 CTGTTGTTATGAATTTCAATCATTTAGTTTAAATCACTATCTGTTTGGTTTATATAAAA 1001
DB |||
QY 915 AGGCTAAATGTTTCTCTCAATTTAGCC-TTTCTTATTTATCTTT 959
DB |||
QY 1002 AGGCTAAATGTTTCTCTCAATTTAGCC-TTTCTTATTTATCTTT 1047
DB |||
```

RESULT 7

AAF25264

ID AAF25264 standard; DNA; 952 BP.

XX

```
AC AAF25264;
XX
DT 30-APR-2001 (first entry)
XX
DE Nucleotide sequence of the DsrA protein from strain CIPA77.
XX
KW DsrA protein; outer membrane protein; serum resistance; vaccine;
KW chancroid disease; genital ulcer disease; ss.
XX
OS Haemophilus ducreyi.
XX
FH Key Location/Qualifiers
FT CDS 135..929
FT FT /*tag= a
FT FT /product= "DsrA protein"
XX
PN WO200104138-A1.
XX
PD 18-JAN-2001.
XX
PF 07-JUL-2000; 2000WO-US018834.
XX
PR 09-JUL-1999; 99US-0143257P.
XX
PA (JUNC-) UNIV NORTH CAROLINA.
PA (ELKI/) ELKINS C.
XX
PI Elkins C;
XX
DR WPI; 2001-138311/14.
DR P-PSDB; AAB31707.
XX
PT Novel purified DsrA protein of Haemophilus ducreyi which confers serum
PT resistance to the bacteria used to produce vaccines that induce immune
PT response against the bacteria in subject at risk of developing chancroid.
XX
PS Claim 1; Page 53-54; 80pp; English.
XX
CC The present sequence encodes a DsrA 30 kDa protein of Haemophilus
CC ducreyi. DsrA is an outer membrane protein that confers serum resistance
CC to the bacteria. DsrA antisense oligonucleotides are useful for detecting
CC a polynucleotide which encodes DsrA in a biological sample. The DsrA
CC polypeptide is used to produce vaccine compositions, which are useful for
CC inducing a protective immune response in a subject at risk of developing
CC Haemophilus ducreyi infection e.g. chancroid, a genital ulcer disease
CC transmitted by sexual contact. DsrA, its catalytic or immunogenic
CC fragments or peptides can be used for screening libraries of compounds in
CC a variety of drug screening techniques. The proteins and peptides may be
CC used as antigens in immunoassays for detection of Haemophilus ducreyi in
CC various tissues and body fluids, e.g. blood, spinal fluid, sputum etc.
CC The nucleic acids are useful for the preparation of DsrA proteins
XX
SQ Sequence 952 BP; 311 A; 146 C; 172 G; 323 T; 0 U; 0 Other;
```

```
Query Match 71.9%; Score 839.6; DB 4; Length 952;
Best Local Similarity 96.7%; Pred. No. 1.7e-172;
Matches 888; Conservative 0; Mismatches 4; Indels 26; Gaps 2;
```

```
QY 1 ATAAATACGTCATTGACATTTTAAATGTAAGGTAGAATAAGAAAGTAAATTTCTATATTT 60
DB |||
QY 40 ATAAATACGTCATTGACATTTTAAATGTAAGGTAGAATAAGAAAGTAAATTTCTATATTT 94
DB |||
QY 61 TACAATCAAGATTTGACAAATTTTAAATGTAAGGTAGAATAAGAAAGTAAATTTCTATATTT 120
DB |||
QY 95 TACAATCAAGATTTGACAAATTTTAAATGTAAGGTAGAATAAGAAAGTAAATTTCTATATTT 154
DB |||
QY 121 TGCCGTAGTGGGATTTAGCTTGTCTTACTATTACAAATGGCTCAGCAGCGCCAAAGTT 180
DB |||
QY 155 TGCCGTAGTGGGATTTAGCTTGTCTTACTATTACAAATGGCTCAGCAGCGCCAAAGTT 214
DB |||
QY 181 TGCTGAGTATCTTCTTTGTATAGCTATGATGATGATGATGATGATGATGATGATGATGATG 240
DB |||
QY 215 TGCTGAGTATCTTCTTTGTATAGCTATGATGATGATGATGATGATGATGATGATGATGATG 274
DB |||
```

241 GTCTAATGAAGCGGTTTCGATATTAAAGTCCAGGATTAAATGAAGCCAAAGAATG 300
 275 GTCTAATGAAGCGGTTTCGATATTAAAGTCCAGGATTAAATGAAGCCAAAGAATG 334
 301 GATTCTTAACAGGCTACTTACTGTGAATTAACAGATTATATGCTTATATCTCTGTCT 360
 335 GATTCTTAACAGGCTACTTACTGTGAATTAACAGATTATATGCTTATATCTCTGTCT 394
 361 CGTCACATATGCTCTCGCGTTCTCTAGCCCTATATCTGTTATATCGATGCTGATCC 420
 395 CGTGACATATGCTCATGAGTCTCTCTAGCTCTATCTATCTGTTATATCGATGCTGATCC 454
 421 TGATCAACTTGGAAATAATCGGCAGCAGCTGAAATTTGAATTTGTATAGTTATTTAAACGA 480
 455 TGATCAACTTGGAAATAATCGGCAGCAGCTGAAATTTGAATTTGTATAGTTATTTAAACGA 514
 481 TTTAAGACACGATTTTAAATTTAAAGTCTTGTGATGCGATTTTCCAAAATAAACAAA 540
 515 TTTAAGACACGATTTTAAATTTAAAGTCTTGTGATGCGATTTTCCAAAATAAACAAA 574
 541 TATTGATCTATAAGTAAATTTTACTAGAACTGGTACTTATTAGATGATTTCTTATCG 600
 575 TATTGATCTATAAGTAAATTTTACTAGAACTGGTACTTATTAGATGATTTCTTATCG 634
 601 TATGATGGAAC-----AAAATACACATAATATCAATAAGTTGTC 639
 635 TATGATGGAACAAATACACATAATATCAATAAGTTGTC 694
 640 TAAAGATTGCAACTGTTTGTAGCAACCAATCAGCATTTGTTAGTTAGTGCACAAA 699
 695 TAAAGATTGCAACTGTTTGTAGCAACCAATCAGCATTTGTTAGTTAGTGCACAAA 754
 700 TGGTGTAGGCAAAACGCGTTTCTGCTGCGGTAGGAGTTATAGATTAATCACTGCATT 759
 755 TGGTGTAGGCAAAACGCGTTTCTGCTGCGGTAGGAGTTATAGATTAATCACTGCATT 814
 760 AGCCATTGGTGTGCGCTCACGCATTACTGATCGCTTTACCGCTAAAGCGGTGTAGCGTT 819
 815 AGCCATTGGTGTGCGCTCACGCATTACTGATCGCTTTACCGCTAAAGCGGTGTAGCGTT 874
 820 CAATACCTTAAATGGCGGATGCTTATGFGTCTTCTGTTGGTTATGAATTTCTAATCATT 879
 875 CAATACCTTAAATGGCGGATGCTTATGFGTCTTCTGTTGGTTATGAATTTCTAATCATT 934
 880 ACGTTTAACTCAATCG 897
 935 ACGTTTAACTCAATCG 952

RESULT 8
 AAF25265
 ID AAF25265 standard; DNA; 899 BP.
 XX AC AAF25265;
 XX DT 30-APR-2001 (first entry)
 XX DE Nucleotide sequence of the DsrA protein from strain CIP542 (Can).
 XX KW DsrA protein; outer membrane protein; serum resistance; vaccine;
 XX KW chancroid disease; genital ulcer disease; ss.
 XX OS Haemophilus ducreyi.
 XX FH Location/Qualifiers
 XX CDS 139..864
 XX FT /*tag= a
 XX FT /product= "DsrA protein"
 XX FT /note= "no termination codon given"
 XX FN W0200104138-A1.
 XX

PD 18-JAN-2001.
 XX 07-JUL-2000; 2000WO-US018834.
 XX 09-JUL-1999; 99US-0143257P.
 XX (UUNC-) UNIV NORTH CAROLINA.
 PA (BLKI/) ELKINS C.
 XX Elkins C;
 XX WPI; 2001-138311/14.
 DR P-PSDB; AAB31708.
 XX Novel purified DsrA protein of Haemophilus ducreyi which confers serum
 PT resistance to the bacteria used to produce vaccines that induce immune
 PT response against the bacteria in subject at risk of developing chancroid.
 XX Claim 1; Page 54; 80pp; English.
 PS
 CC The present sequence encodes a DsrA 30 kDa protein of Haemophilus
 CC ducreyi. DsrA is an outer membrane protein that confers serum resistance
 CC to the bacteria. DsrA antisense oligonucleotides are useful for detecting
 CC a polynucleotide which encodes DsrA in a biological sample. The DsrA
 CC polypeptide is used to produce vaccine compositions, which are useful for
 CC inducing a protective immune response in a subject at risk of developing
 CC Haemophilus ducreyi infection e.g. chancroid, a genital ulcer disease
 CC transmitted by sexual contact. DsrA, its catalytic or immunogenic
 CC fragments or peptides can be used for screening libraries of compounds in
 CC a variety of drug screening techniques. The proteins and peptides may be
 CC used as antigens in immunoassays for detection of Haemophilus ducreyi in
 CC various tissues and body fluids, e.g. blood, spinal fluid, sputum etc. .
 CC The nucleic acids are useful for the preparation of DsrA proteins
 XX
 SQ Sequence 899 BP; 291 A; 136 C; 163 G; 309 T; 0 U; 0 Other;
 Query Match 71.0%; Score 829.2; DB 4; Length 899;
 Best Local Similarity 99.6%; Pred. No. 3e-170;
 Matches 831; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 ATAAATACGTCATGTGACATTTTAAATGTAAGTAGAATAAGAAAGTAATTCATATT 60
 DB 39 ATAAATACGTCATGTGACATTTTAAATGTAAGTAGAATAAGAAAGTAATTCATATT 98
 QY 61 TACAATCAAGATTGACAAATTTTAAATGTAAGTAGAATAAGAAAGTAATTCATATT 120
 DB 99 TACAATCAAGATTGACAAATTTTAAATGTAAGTAGAATAAGAAAGTAATTCATATT 158
 QY 121 TGCCGTAGTGGGATTAGCTTTGTTCTACTATTACAACAATGGCTCAGCGCGCAAGTT 180
 DB 159 TGCCGTAGTGGGATTAGCTTTGTTCTACTATTACAACAATGGCTCAGCGCGCAAGTT 218
 QY 181 TGCTGGAGTATCTTTGTTGATAGCTATGATGATGTAAGGTAAATCGACTTG 240
 DB 219 TGCTGGAGTATCTTTGTTGATAGCTATGATGATGTAAGGTAAATCGACTTG 278
 QY 241 GTCTAATGAAGCGGTTTCGATATTAAAGTCCAGGATTAAATGAAGCCAAAGAATG 300
 DB 279 GTCTAATGAAGCGGTTTCGATATTAAAGTCCAGGATTAAATGAAGCCAAAGAATG 338
 QY 301 GATTCTTAACAGGCTACTTACTGTGAATTTAAAGTCCAGGATTAAATGAAGCCAAAGAATG 360
 DB 339 GATTCTTAACAGGCTACTTACTGTGAATTTAAAGTCCAGGATTAAATGAAGCCAAAGAATG 398
 QY 361 CGTGACATATGCTCTCGCGTTCTCTAGCCCTATATCTGTTATATCGATGCTGATCC 420
 DB 399 CGTGACATATGCTCTCGCGTTCTCTAGCCCTATATCTGTTATATCGATGCTGATCC 458
 QY 421 TGATCAACTTGGAAATAATCGGCAGCAGCTGAAATTTGAATTTGTATAGTTATTTAAACGA 480
 DB 459 TGATCAACTTGGAAATAATCGGCAGCAGCTGAAATTTGAATTTGTATAGTTATTTAAACGA 518
 QY 481 TTTAAGACACGATTTTAAATTTAAAGTCTTGTGATGCGATTTTCCAAAATAAACAAA 540

```

Db 519 TTTAAGACACGATTTTAAATTAAGTTCTTGATGACGATTTTCCAAAATAAACAAA 578
Qy 541 TATTGATACATAAGTAAATTTTACAGAACTGGGTACTTATTTAGATGATCTTATCG 600
Db 579 TATTGATACATAAGTAAATTTTACAGAACTGGGTACTTATTTAGATGATCTTATCG 638
Qy 601 TATGATGAAACAAATACACATAATATCAATAAGTCTCTAAAGAAATTCGAAACTGGTTT 660
Db 639 TATGATGAAACAAATACACATAATATCAATAAGTCTCTAAAGAAATTCGAAACTGGTTT 698
Qy 661 AGCCAAACCAATCAGCATTTGCTATGTTAGTGCAACCAATAGGTGTAGGCAAAAGAGCGT 720
Db 699 AGCCAAACCAATCAGCATTTGCTATGTTAGTGCAACCAATAGGTGTAGGCAAAAGAGCGT 758
Qy 721 TTCTGCTGCGGTAGGAGGTTATAGAGATAAACTGCAATTAGCCATTTGGTGTGCGCTCAG 780
Db 759 TTCTGCTGCGGTAGGAGGTTATAGAGATAAACTGCAATTAGCCATTTGGTGTGCGCTCAG 818
Qy 781 CATTACTGATCGCTTTTACCCTAAAGCGGGTGTAGCGTTTCAATACCTACCAATGG 834
Db 819 CATTACTGATCGCTTTTACCCTAAAGCGGGTGTAGCGTTTCAATACCTTCTATCG 872

```

RESULT 9

AAF25267 standard; DNA; 923 BP.

XX AAF25267;

XX 30-APR-2001 (first entry)

XX Nucleotide sequence of the DsrA protein from strain CH1A.

XX DsrA protein; outer membrane protein; serum resistance; vaccine;
 XX chancroid disease; genital ulcer disease; ss.

XX Haemophilus ducreyi.

XX Key Location/Qualifiers
 XX CDS 45..836

XX FT /*tag= a
 XX FT /product= "DsrA protein"

XX WO200104138-A1.

XX 18-JAN-2001.

XX 07-JUL-2000; 2000WO-US018834.

XX 09-JUL-1999; 99US-0143257P.

XX (UNNC-) UNIV NORTH CAROLINA.

XX (ELKI/) ELKINS C.

XX Elkins C;

XX WPI; 2001-138311/14.
 XX P-PSDB; AAB31710.

XX Novel purified DsrA protein of Haemophilus ducreyi which confers serum
 XX resistance to the bacteria used to produce vaccines that induce immune
 XX response against the bacteria in subject at risk of developing chancroid.

XX Claim 1; Page 56; 80pp; English.

XX The present sequence encodes a DsrA 30 kDa protein of Haemophilus
 XX ducreyi. DsrA is an outer membrane protein that confers serum resistance
 XX to the bacteria. DsrA antisense oligonucleotides are useful for detecting
 XX a polynucleotide which encodes DsrA in a biological sample. The DsrA
 XX polypeptide is used to produce vaccine compositions, which are useful for
 XX inducing a protective immune response in a subject at risk of developing
 XX Haemophilus ducreyi infection e.g. chancroid, a genital ulcer disease

CC transmitted by sexual contact. DsrA, its catalytic or immunogenic
 CC fragments or peptides can be used for screening libraries of compounds in
 CC a variety of drug screening techniques. The proteins and peptides may be
 CC used as antigens in immunoassays for detection of Haemophilus ducreyi in
 CC various tissues and body fluids, e.g. blood, spinal fluid, sputum etc. .
 CC The nucleic acids are useful for the preparation of DsrA proteins

XX
 SQ Sequence 923 BP; 291 A; 151 C; 174 G; 307 T; 0 U; 0 Other;

Query Match 70.0%; Score 818; DB 4; Length 923;
 Best Local Similarity 94.7%; Pred. No. 8.1e-168;
 Matches 877; Conservative 0; Mismatches 25; Indels 24; Gaps 2;

```

Qy 57 TATTTACAATCAAGATTGACAAATTTATTTACTTAATGAGGTGATTATGAAATTTAAATGTT 116
Db 1 TATTTACAATCAAGATTGACAAATTTATTTACTTAATGAGGTGATTATGAAATTTAAATGTT 60
Qy 117 TAGTTGCCGTAGTGGGATTAGCTTTGTTTACTACTATTACAACAATGGCTCAGCAGCGCCAA 176
Db 61 TAGTTGCCGTAGTGGGATTAGCTTTGTTTACTACTATTACAACAATGGCTCAGCAGCGCCAA 120
Qy 177 AGTTTGGCTGGAGTATCTTCTTTGTATAGCTATGAGTATGACTATGTAAGGGTAAATGGA 236
Db 121 AGTTTGGCTGGAGTATCTTCTTTGGATAGCTATGAGTATGACTATGTAAGGGTAAATGGA 180
Qy 237 CTTTGGTCTAATGAAGCGGTTTCGATATTAAAGTGCAGGGATTAAATCAAGCCAAAG 296
Db 181 CTTTGGTCTGAAAAGACCGGTTTCGATATTAAAGCGCAGGGATTAAATCAAGCCAAAG 240
Qy 297 AATGGATTTCTAAACAGGCTACTTATCTTGAATTTACAGCAATTATATGCCCTTATACCTCG 356
Db 241 AATGGATTTCTAGACAGGCTACTTATCTTGGATTACAGCAATTATATGCCCTTATACCTCG 300
Qy 357 TTCTCGTGCATATGCTCTCGCGTTTCTCTAGCCCTATCTACTGTTATATCCGATGCTG 416
Db 301 TTCTCGTGCATATGCTCTCTGCAG---AACCTAACACTGTACTGTTTATATCCGATGCTG 357
Qy 417 ATCCTGATCAACTTGGAAATAAATCGGCAGCAGCTGAAATTTGAATTTGTATAGTTATTTTA 476
Db 358 ATCCTGATCAACTTGGAAATAAATCGGCAGCAGCTGAAATTTGAATTTGTATAGTTATTTTA 417
Qy 477 ACGATTTTAAGACACAGATTTTAAATTTAAAGTTCTTGTATGACACGTATTTCCAAAATAAAC 536
Db 418 ACGATTTTAAGACACAGGTTTAAATTTAAATGTTCTTGTATGACACGTATTTCCAAAATAAAC 477
Qy 537 AAAAATTTGATCTATAAGTAAATTTTACTAGAACTGGGTACTTATTTAGATGATTCCT 596
Db 478 AAAAATTTGATCTATAAGTAAATTTTACTAAAACTGGGTACTTATTTAGATGATTCCT 537
Qy 597 ATCGTATGATGGAAC-----AAAATACACATATATCAATAAGT 635
Db 538 ATCGTATGATGGAACAAAATAACATATATCAATAATAACATAATATCAATAAGT 597
Qy 636 TGTCTAAAGAAATTCGAAACTGGTTTAGCCAAACCAATCAGCATTGTCTATGTTAGTGCAAC 695
Db 598 TGTCTAAAGAAATTCGAAACTGGTTTAGCCAAACCAATCAGCATTGTCTATGTTAGTGCAAC 657
Qy 696 CAAATCGTGTAGGCAAAACGAGCGTTTCTGCTCGGTAGGAGGTTATAGAGATAAAACTG 755
Db 658 CAAATCGTGTAGGCAAAACGAGCGTTTCTGCTCGGTAGGAGGTTATAGAGATAAAACTG 717
Qy 756 CATTAGCCATTGTTGTCGGCTCAGCATTACTGATCGCTTTACCGCTTAAGCGGGTGTAG 815
Db 718 CATTAGCCATTGTTGTCGGCTCAGCATTACTGATCGCTTTACCGCTTAAGCGGGTGTAG 777
Qy 816 CGTTCAATACCTACAAATGCGGCATGCTTTATGCTGCTTCTGTTGTTATGAATTCCTAAT 875
Db 778 CGTTCAATACCTACAAATGCGGCATGCTTTATGCTGCTTCTGTTGTTATGAATTCCTAAT 837
Qy 876 CATTAGCTTTAATCAGTAAATCGTTTTTGGTTTATATAAAGGCTAAATGTTTCTCCTCAC 935
Db 838 CATTAGCTTTAATCAGTAAATCGTTTTTGGTTTATATAAAGGCTAAATGTTTCTCCTCAC 897

```

Qy 936 ATTTAGCCTTCTTATTATCTTTGT 961
|||||
Db 898 ATTTAGCCTTCTTATTATCTTTGT 923
|||||

RESULT 10

AAC79587/c

ID AAC79587 standard; DNA; 4426 BP.

XX

AC AAC79587;

XX

DT 08-FEB-2001 (first entry)

XX

DE Virulence gene #7.

XX

KW Virulence gene; antibacterial; vaccine; bacterial infection; septicemia;
bronchopneumonia; rhinitis; wound infection; ss.

XX

OS Pasteurella multocida.

XX

PN WO200061724-A2.

XX

PD 19-OCT-2000.

XX

PF 06-APR-2000; 2000WO-US009218.

XX

PR 09-APR-1999; 99US-0128689P.

XX

PT 10-SEP-1999; 99US-0153453P.

XX

PA (PHAA) PHARMACIA & UPJOHN INC.

XX

PI Lowery DE, Fuller TE, Kennedy MJ;

XX

DR WPI; 2000-647422/62.

XX

DR P-PSDB; AAB44527.

XX

PT Attenuated Pasteurellaceae bacteria comprising mutations in virulence
genes, useful as a live attenuated vaccine against bacterial infections.

XX

PS Claim 7; Page 83-85; 32pp; English.

XX

CC The family Pasteurellaceae encompasses several pathogens that infect a
wide variety of animals. The present invention relates to virulence genes
from Pasteurellaceae. The present sequence is one such virulence gene.
The present sequence may be mutated in order to produce an inactive gene.
The inactive virulence gene may in turn be used to produce a vaccine,
which is useful for treating bacterial infections such as septicemias,
bronchopneumonias, rhinitis and wound infections

XX

SQ Sequence 4426 BP; 1343 A; 882 C; 827 G; 1374 T; 0 U; 0 Other;

Query Match

Best Local Similarity 5.8%; Score 67.2; DB 3; Length 4426;

Matches 90; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

Qy 1038 ACTTTCACCTTTTTCACCATATTATCGCCACTTCTAAACGGTAATTAAGTTGGTTT 1097
|||||Db 2512 ACTTTGGCTTTTAAATCAATGTTGTCATTCCTCTAGATCGTGATTTAGCCCAITA 2453
|||||Qy 1098 AGCCTAAATGGGTACCTTCTATCGGAATTTTCTAAATGTTCTAAATTAAGCCGTTA 1157
|||||Db 2452 ATTTACATACCGTTCCTTCTCTGGAATTTCTTCTAAATGCTCTAAATTAAGCCGTTG 2393
|||||Qy 1158 AAGTGGC 1165
|||||

Db 2392 AAGGTACG 2385

RESULT 11

ABQ83464/c

ID ABQ83464 standard; DNA; 4426 BP.

XX

AC ABQ83464;

XX

DT 07-AUG-2003 (revised)

XX

DT 24-JAN-2003 (first entry)

XX

DE Pasteurella multocida exbB gene SEQ ID NO:13.

XX

KW Antibacterial; vaccine; gram negative bacterial virulence gene;
identification; virulence; Pasteurellaceae; gene; ds.

XX

OS Pasteurella multocida.

XX

PN WO200275507-A2.

XX

PD 26-SEP-2002.

XX

PF 17-JAN-2002; 2002WO-US001971.

XX

PR 15-MAR-2001; 2001US-00809665.

XX

PA (PHAA) PHARMACIA & UPJOHN CO.

XX

PI Lowery DE, Fuller TE, Kennedy MJ;

XX

DR WPI; 2002-740868/80.

XX

DR P-PSDB; ABP54479.

XX

PT New mutant gram-negative bacteria, useful as vaccines and for identifying
new anti-bacterial agents that target virulence genes and their products.

XX

PS Claim 5; Page 87-90; 350pp; English.

XX

CC The present invention describes a gram-negative bacteria comprising a
mutation in a gene, where the mutation results in decreased activity of a
gene product encoded by the mutated gene. Also described is a method for
producing a gram-negative bacteria mutant or an attenuated
Pasteurellaceae bacteria. The mutated genes have antibacterial activity
and can be used in vaccines. The gram-negative bacteria or the attenuated
Pasteurellaceae bacteria can be used as vaccines in the fields of human
medicine or veterinary medicine, and for identifying new antibacterial
agents that target the virulence genes and their products. ABQ83458 to
ABQ83578 and ABP54473 to ABP54551 represents sequences used in the
CC exemplification of the present invention. (Updated on 07-AUG-2003 to
correct OS field.)

XX

SQ Sequence 4426 BP; 1342 A; 882 C; 827 G; 1374 T; 0 U; 1 Other;

Query Match

Best Local Similarity 5.8%; Score 67.2; DB 6; Length 4426;

Matches 90; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

Qy 1038 ACTTTCACCTTTTTCACCATATTATCGCCACTTCTAAACGGTAATTAAGTTGGTTT 1097
|||||Db 2512 ACTTTGGCTTTTAAATCAATGTTGTCATTCCTCTAGATCGTGATTTAGCCCAITA 2453
|||||Qy 1098 AGCCTAAATGGGTACCTTCTATCGGAATTTTCTAAATGTTCTAAATTAAGCCGTTA 1157
|||||Db 2452 ATTTACATACCGTTCCTTCTCTGGAATTTCTTCTAAATGCTCTAAATTAAGCCGTTG 2393
|||||Qy 1158 AAGTGGC 1165
|||||

Db 2392 AAGGTACG 2385

RESULT 12

AAZ54498

ID AAZ54498 standard; DNA; 1095 BP.

XX

AC AAZ54498;

XX

DT 21-MAR-2000 (first entry)

XX

DE Neisseria meningitidis ORF 961 partial DNA sequence SEQ ID NO:2943.

XX

KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
KW antibacterial; gene therapy; ds.
XX
OS Neisseria meningitidis.
XX
PN WO9957280-A2.
XX
XX
XX
PD 11-NOV-1999.
XX
XX
XX
PF 30-APR-1999; 99WO-US009346.
XX
XX 01-MAY-1998; 98US-0083758P.
PR 31-JUL-1998; 98US-0094869P.
PR 02-SEP-1998; 98US-0088994P.
PR 02-SEP-1998; 98US-0099062P.
PR 09-OCT-1998; 98US-0103749P.
PR 09-OCT-1998; 98US-0103794P.
PR 09-OCT-1998; 98US-0103796P.
PR 25-FEB-1999; 99US-0121528P.
XX
XX (CHIR) CHIRON CORP.
PA (GENO-) INST GENOMIC RES.
XX
XX Fraser C, Galeotti C, Grandi G, Hickey E, Maignani V, Mora M;
PI Petersen J, Pizzuoli R, Ratti G, Scalato E, Scarselli M;
PI Tettelin H, Venter JC;
XX
DR WPI; 2000-062150/05.
DR P-PSDB; AAY75736.
XX
XX Novel Neisserial polypeptides predicted to be useful antigens for
PT vaccines and diagnostics.
XX
XX Claim 7; Page 1377; 1453pp; English.
XX
XX AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
CC represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides
CC and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ5473 represent
CC PCR primers used in the exemplification of the present invention. The
CC polypeptides, the polynucleotides, antibodies and compositions of the
CC invention can be used as vaccines, as diagnostic reagents, and as
CC immunogenic compositions. The polypeptides can be used in the manufacture
CC of medicaments for treating or preventing infection due to Neisserial
CC bacteria (e.g. meningitis and septicemia), to detect the presence of
CC Neisseria bacteria, or to raise antibodies. They may also be used to
CC screen for agonists or antagonists, which may themselves have use as
CC antibacterial agents. The polynucleotides of the invention may also be
CC used in gene therapy protocols
XX
SQ Sequence 1095 BP; 363 A; 285 C; 249 G; 198 T; 0 U; 0 Other;
Query Match 5.1%; Score 59.8; DB 3; Length 1095;
Best Local Similarity 55.6%; Pred. No. 0.0023;
Matches 115; Conservative 0; Mismatches 92; Indels 0; Gaps 0;
QY 621 ATAATATCAATAGTCTCTAAGAATGCAAACTGGTTAGCCAACTACGATCAGCATGT 680
DB 839 AAAACGTAGCTAATCTCGCAAGAAACCGCCCAAGCGCTTGAGAACAAAGCGCGCTCT 898
QY 681 CTATGTTAGTGCACCAATGGTGTAGGCAAAACGACGCTTCTGTCGGTAGGAGTT 740
DB 899 CCGGCCCTGTTCCAACTTCAACAGTGGTGGTTCATGTAAACGGCTGCAGTCGGCGGCT 958
QY 741 ATAGAGATAAACTGCATTAGCATTGCTGCGCTCAACGATTAATCTGCTTTACCG 800
DB 959 ACAATCCGATCGGAGTGCCTATCGTACCGGCTTCGGCTTACCGAACTTTCGGC 1018
QY 801 CTAAGCGGGTGTAGCGTTCAATACCT 827
DB 1019 CCAAGCAGCGGTGGCAGTCGGCACTT 1045

RESULT 13
AA81472/C
ID AAA81472 standard; DNA; 16526 BP.
XX
XX AC AA81472;
XX
XX 04-DEC-2000 (first entry)
XX
XX N. meningitidis partial DNA sequence gnm_20 SEQ ID NO:20.
XX
XX Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;
KW antigen; vaccine; diagnosis; infection; antibacterial; identification;
KW Meningococcus B; MenB; ds.
XX
XX Neisseria meningitidis.
OS
XX WO200022430-A2.
PN
XX 20-APR-2000.
XX
XX 08-OCT-1999; 99WO-US023573.
PF
XX 09-OCT-1998; 98US-0103794P.
PR 30-APR-1999; 99US-0132068P.
XX
XX (CHIR) CHIRON CORP.
XX
XX Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC;
PI Maignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scariato V;
PI Rappuoli R, Pizzuoli R;
XX
XX WPI; 2000-318079/27.
DR
XX Isolated nucleotide sequences of Neisseria meningitidis which can be used
PT in the diagnosis and treatment of N. meningitidis infection and other
PT Neisserial infections, for example, N. gonorrhoea.
XX
XX Claim 7; Page 466-471; 1760pp; English.
XX
XX The present invention describes methods of obtaining immunogenic proteins
CC from Neisseria genomic sequences. AA81453 to AA82414 represent
CC specifically claimed Neisseria meningitidis genomic DNA sequences;
CC AA81260 to AA81303 and AA825620 to AA825663 represent Neisseria DNA
CC sequences and their corresponding proteins; AA81254 to AA81259 and
CC AA81304 to AA81321 represent PCR primers used in the isolation of
CC Neisseria meningitidis DNA sequences; and AA81322 to AA81452 represent
CC Neisseria meningitidis MenB polynucleotide ORF sequences, which are all
CC used in the exemplification of the present invention. The nucleic acid
CC sequences, protein sequences, and antibodies against them, can be used in
CC the manufacture of a composition. The composition can be used as a
CC medicament (or in the manufacture of a medicament) for treating,
CC preventing or diagnosing infection due to Neisserial bacteria. For
CC example, some of the identified proteins could be components of vaccines
CC against Meningococcus B; against all serotypes; and/or against all
CC pathogenic Neisseriae. Identification of sequences from the bacterium
CC will also facilitate production of biological probes, particularly
CC organism-specific probes. Attempts to make efficacious Meningococcus B
CC vaccines have failed mainly due to antigen tolerance. Multivalent
CC vaccines have also been tried but none have successfully overcome
CC antigenic variability. The provision of further, complete sequences may
CC provide an opportunity to identify secreted or surface exposed proteins
CC that may be presumed targets for the immune system and which are not
CC antigenically variable or at least more conserved than other more
XX variable regions
SQ Sequence 16526 BP; 3183 A; 4006 C; 5067 G; 4268 T; 0 U; 2 Other;
Query Match 5.1%; Score 59.8; DB 3; Length 16526;
Best Local Similarity 55.6%; Pred. No. 0.0042;
Matches 115; Conservative 0; Mismatches 92; Indels 0; Gaps 0;
QY 621 ATAATATCAATAGTCTCTAAGAATGCAAACTGGTTAGCCAACTACGATCAGCATGT 680

Db 14258 AAAACGTAGCTAATCTCGGCAAGAACAACC CGCCAAAGCCTTGCAGAACAAAGCCGGCTCT 14199

Qy 681 CTATGTTTAGTGCAACCAAATAGTGTAGGCCAAAACAGAGCGTTTTCTGCTGCGGTAGGAGGTT 740

Db 14198 CCGGCCTGTTCCAACACTTACAACTGTCGGTTCGAATTGTAACGGCTGCAGTCGGCGGCT 14139

Qy 741 ATGAGAGATAAACTGCANTGATGCGCATTCGGCTCACGCATTACTGATCGCTTTACCG 800

Db 14138 ACAAAATCCGAATCGGCAGTCCCATCGTACCGGCTTCCGCTTTACCGAAAACTTTGCCG 14079

Qy 801 CTAAGCGGCTGTAGCGTTCAATACCT 827

Db 14078 CCAAAAGCAGCGTGGCAGTCGGCACTT 14052

RESULT 14
Continuation (7 of 9) of AA81489 from base 600001 (N. meningitidis partial DNA sequence)
WP Sequence split into 9 fragments LOCUS AA81489 Accession Aaa81489
Fragment Name Begin End
WP AA81489_0 1 110000
WP AA81489_1 100001 210000
WP AA81489_2 200001 310000
WP AA81489_3 300001 410000
WP AA81489_4 400001 510000
WP AA81489_5 500001 610000
WP AA81489_6 600001 710000
WP AA81489_7 700001 810000
WP AA81489_8 800001 837096

Query Match 5.1%; Score 59.8; DB 3; Length 110000;
Best Local Similarity 55.6%; Pred. No. 0.0063;
Matches 115; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

Qy 621 ATAATATCAATAGTTGTTCTTAAGAATGCAAACTGGTTTAGCCAACTATCAGCATTTGT 680

Db 65951 AAAACGTAGCTAATCTCGCAAGAAACCCGCCAAGGCTTGCAGAACAAAGCCGGCTCT 66010

Qy 681 CTATGTTTAGTGCAACCAAATAGTGTAGGCCAAAACAGAGCGTTTTCTGCTGCGGTAGGAGGTT 740

Db 66011 CCGGCCTGTTCCAACCTTTACAAGTCGGTTCGAATGTAACGGCTGCAGTCGGCGGCT 66070

Qy 741 ATGAGAGATAAACTGCANTGATGCGCATTCGGCTCACGCATTACTGATCGCTTTACCG 800

Db 66071 ACAATCCGAATCGGCAGTCCCATCGTACCGGCTTCCGCTTTACCGAAAACTTTGCCG 66130

Qy 801 CTAAGCGGCTGTAGCGTTCAATACCT 827

Db 66131 CCAAGCAGCGTGGCAGTCGGCACTT 66157

RESULT 15
AAF21613
ID AAF21613 standard; DNA; 172325 BP.
XX AC
XX AAF21613;
XX AC
DT DT
DE DE
DE DE
XX Neisseria meningitidis B nucleotide sequence SEQ ID NO:114.
XX Neisseria meningitidis; Neisseria gonorrhoeae; immunogenic; vaccine;
KW diagnosis; antigen; detection; infection; gene therapy; antibacterial;
KM ds.
XX OS
OS Neisseria meningitidis.
XX OS
XX WO200066791-Al.
XX PN
XX PD
PD 09-NOV-2000.
XX PF
PF 08-MAR-2000; 2000WO-US005928.
XX XX

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model
Run on: May 15, 2004, 01:33:10 ; Search time 4881 Seconds
(without alignments)
10371.781 Million cell updates/sec

Title: US-10-030-529A-1
Perfect score: 1168
Sequence: 1 atcaatagctcattgacatt.....aagccgttaaaggtagggac 1168

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0
Searched: 3470272 seqs, 21671516995 residues
Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 10%
Maximum Match 100%
Listing first 45 summaries

Database :	1:	GenBank:	2:	gb_hgtg:	3:	gb_in:	4:	gb_om:	5:	gb_ov:	6:	gb_pat:	7:	gb_ph:	8:	gb_pl:	9:	gb_pr:	10:	gb_ro:	11:	gb_sta:	12:	gb_sy:	13:	gb_un:	14:	gb_vi:	15:	em_ba:	16:	em_fun:	17:	em_hum:	18:	em_in:	19:	em_mu:	20:	em_om:	21:	em_or:	22:	em_ov:	23:	em_pat:	24:	em_ph:	25:	em_pl:	26:	em_ro:	27:	em_sta:	28:	em_un:	29:	em_vi:	30:	em_htg_hum:	31:	em_htg_inv:	32:	em_htg_other:	33:	em_htg_mus:	34:	em_htg_pln:	35:	em_htg_rod:	36:	em_htg_mam:	37:	em_htg_vrt:	38:	em_sy:	39:	em_htgo_hum:	40:	em_htgo_mus:	41:	em_htgo_other:
	1:	gb_ba:	2:	gb_hgtg:	3:	gb_in:	4:	gb_om:	5:	gb_ov:	6:	gb_pat:	7:	gb_ph:	8:	gb_pl:	9:	gb_pr:	10:	gb_ro:	11:	gb_sta:	12:	gb_sy:	13:	gb_un:	14:	gb_vi:	15:	em_ba:	16:	em_fun:	17:	em_hum:	18:	em_in:	19:	em_mu:	20:	em_om:	21:	em_or:	22:	em_ov:	23:	em_pat:	24:	em_ph:	25:	em_pl:	26:	em_ro:	27:	em_sta:	28:	em_un:	29:	em_vi:	30:	em_htg_hum:	31:	em_htg_inv:	32:	em_htg_other:	33:	em_htg_mus:	34:	em_htg_pln:	35:	em_htg_rod:	36:	em_htg_mam:	37:	em_htg_vrt:	38:	em_sy:	39:	em_htgo_hum:	40:	em_htgo_mus:	41:	em_htgo_other:
	1:	gb_ba:	2:	gb_hgtg:	3:	gb_in:	4:	gb_om:	5:	gb_ov:	6:	gb_pat:	7:	gb_ph:	8:	gb_pl:	9:	gb_pr:	10:	gb_ro:	11:	gb_sta:	12:	gb_sy:	13:	gb_un:	14:	gb_vi:	15:	em_ba:	16:	em_fun:	17:	em_hum:	18:	em_in:	19:	em_mu:	20:	em_om:	21:	em_or:	22:	em_ov:	23:	em_pat:	24:	em_ph:	25:	em_pl:	26:	em_ro:	27:	em_sta:	28:	em_un:	29:	em_vi:	30:	em_htg_hum:	31:	em_htg_inv:	32:	em_htg_other:	33:	em_htg_mus:	34:	em_htg_pln:	35:	em_htg_rod:	36:	em_htg_mam:	37:	em_htg_vrt:	38:	em_sy:	39:	em_htgo_hum:	40:	em_htgo_mus:	41:	em_htgo_other:

ALIGNMENTS

RESULT 1	AE017152	301191 bp	DNA	linear	BCT 15-AUG-2003
LOCUS	AE017152	Haemophilus ducreyi strain 3500HP	section 2 of 6 of the complete genome.		
DEFINITION	AE017152	Haemophilus ducreyi strain 3500HP			
ACCESSION	AE017152	AE017143			
VERSION	AE017152.1	GI:33147833			
KEYWORDS					
SOURCE		Haemophilus ducreyi 3500HP			
ORGANISM		Haemophilus ducreyi 3500HP			
		Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;			
		Pasteurellaceae; Haemophilus.			
REFERENCE		1 (bases 1 to 301191)			
AUTHORS		Munson, R.S. Jr., Ray, W.C., Mahairas, G., Sabo, P., Mungur, R., Johnson, L., Nguyen, D., Wang, J., Forst, C. and Hood, L.			

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1168	100.0	301191	1	AE017152	AE017152 Haemophil
2	1000	85.6	1000	1	AF187001	AF187001 Haemophil
3	874	74.8	874	1	AF187005	AF187005 Haemophil
4	874	74.8	874	1	AF187007	AF187007 Haemophil
5	829.2	71.0	847	1	AF187004	AF187004 Haemophil
6	816.6	69.9	890	1	AF187002	AF187002 Haemophil
7	816.6	69.9	890	1	AF187003	AF187003 Haemophil
8	814	69.7	916	1	AF187009	AF187009 Haemophil
9	792.8	67.9	923	1	AF187008	AF187008 Haemophil
10	780.6	66.8	894	1	AF187006	AF187006 Haemophil
11	84.6	7.2	1967	1	AB064943	AB064943 Actinobac
12	84.6	7.2	2579	1	AF316502	AF316502 Actinobac
13	68.6	5.9	2216	1	YPYOPA	X13883 Yersinia ps
14	68.4	5.9	2216	1	YPYOPAF	X13880 Yersinia pe
15	68.4	5.9	70305	1	YPCD1	AL117189 Yersinia
16	68.4	5.9	70504	1	AF053946	AF053946 Yersinia
17	68.4	5.9	70559	1	AF074612	AF074612 Yersinia
18	68	5.8	66591	1	AY150843	AY150843 Yersinia
19	67.2	5.8	4426	6	BD252123	BD252123 Anti-Bact
20	67.2	5.8	4426	6	AX553679	AX553679 Sequence
21	67.2	5.8	11632	1	AE006158	AE006158 Pasteurel
22	66.4	5.7	2186	1	YEQOPA80	X13881 Yersinia en
23	66.4	5.7	67720	1	AF336309	AF336309 Yersinia
24	63.8	5.5	2551	1	YEQOPA64	X13882 Y. enterocol
25	61.8	5.3	3167	1	AF399847	AF399847 Escherich
26	60.2	5.2	11279	7	AF151091	AF151091 Prophage
27	60	5.1	3795	7	AF151675	AF151675 Bacteriop
28	60	5.1	9689	7	AF151674	AF151674 Bacteriop
29	59.8	5.1	1760	1	AF452479	AF452479 Neisseria
30	59.8	5.1	1761	1	AF452481	AF452481 Neisseria
31	59.8	5.1	1765	1	AF452477	AF452477 Neisseria
32	59.8	5.1	1767	1	AF452472	AF452472 Neisseria
33	59.8	5.1	1778	1	AF452486	AF452486 Neisseria
34	59.8	5.1	1779	1	AF452466	AF452466 Neisseria
35	59.8	5.1	1782	1	AF452467	AF452467 Neisseria
36	59.8	5.1	1789	1	AF452478	AF452478 Neisseria
37	59.8	5.1	1791	1	AF452485	AF452485 Neisseria
38	59.8	5.1	1794	1	AF452475	AF452475 Neisseria
39	59.8	5.1	9977	1	AE002548	AE002548 Neisseria
40	59.8	5.1	172325	6	AX044035	AX044035 Sequence
41	59.8	5.1	349980	6	AX044034	AX044034 Sequence
42	59.6	5.1	10303	1	AE006108	AE006108 Pasteurel
43	59.2	5.1	972	6	AX698453	AX698453 Sequence
44	59.2	5.1	1870	1	AF452484	AF452484 Neisseria
45	59.2	5.1	1874	1	AF452476	AF452476 Neisseria

TITLE JOURNAL REFERENCE AUTHORS	The Complete Genome Sequence of Haemophilus ducreyi 2 (bases 1 to 301191) Munson, R. S., Jr., Ray, W. C., Mahairas, G., Sabo, P., Mungur, R., Johnson, L., Nguyen, D., Wang, J., Forst, C. and Hood, L.	
	Direct Submission Submitted (04-JUN-2003) Pediatrics, Columbus Children's Research Institute and The Ohio State University, 700 Children's Drive, Columbus, OH 43205, USA	
FEATURES source	Location/Qualifiers 1. .301191 /organism="Haemophilus ducreyi 35000Hp" /mol_type="genomic DNA" /strain="35000HP" /db_xref="taxon:233412" 115. .1143 /genes="apbE" 115. .1143 /locus tag="HD0386" /gene="apbE" /locus_tag="HD0386" /codon_start=1 /transl_table=11 /product="thiamine biosynthesis lipoprotein" /protein_id="AAP95355.1" /db_xref="GI:33147834" /translation="MLVCVASLSFILTACNKSAEQITLOGKTMGTYTYVKYIDDGKIA NLPTAQVKQLDQVLKVNDEMSITQLDSQISRFNKREVNTPFEISPDFATVAAEA IRLNKITEGALDVTGFLVGPDRDLDPQSAAQINQRAAFVIGIDKLAINTBEHK PTUTKSPYNTLAKGFGVDKIADQLEMLGLENLYLEVGEURGKGNLQGLNW QIAIEKPNPEOGQTVITVPLNHLGMATSGNYRNPEDQGNRLSHIIDPKALKPWNH KLASITVLAPTSMTADGLSTGLFVLGAERKALEVAEREKLAIFLIINKGDPGPEQMSQQ FKLLIDQK"	
	gene	1155. .1403 /locus tag="HD0387" 1155. .1403 /locus tag="HD0387" /codon_start=1 /transl_table=11 /product="conserved hypothetical protein" /protein_id="AAP95356.1" /db_xref="GI:33147835" /translation="METLLTFFGFLVIFAMSIGFIIRKGIKSGCGGITALGMKKQM CDCEPCDNLKTKIMEGKADPAKIAQFKKXVASFYEVK"
gene	1155. .1403 /locus tag="HD0388" 1155. .1403 /locus tag="HD0388" /codon_start=1 /transl_table=11 /product="ton-dependent heme receptor A" /protein_id="AAP95357.1" /db_xref="GI:33147836"	
	CDS	1155. .1403 /locus tag="HD0389" 1155. .1403 /locus tag="HD0389" /gene="trmU"
TITLE JOURNAL REFERENCE AUTHORS	The Complete Genome Sequence of Haemophilus ducreyi 2 (bases 1 to 301191) Munson, R. S., Jr., Ray, W. C., Mahairas, G., Sabo, P., Mungur, R., Johnson, L., Nguyen, D., Wang, J., Forst, C. and Hood, L.	
	Direct Submission Submitted (04-JUN-2003) Pediatrics, Columbus Children's Research Institute and The Ohio State University, 700 Children's Drive, Columbus, OH 43205, USA	
FEATURES source	Location/Qualifiers 1. .301191 /organism="Haemophilus ducreyi 35000Hp" /mol_type="genomic DNA" /strain="35000HP" /db_xref="taxon:233412" 115. .1143 /genes="apbE" 115. .1143 /locus tag="HD0386" /gene="apbE" /locus_tag="HD0386" /codon_start=1 /transl_table=11 /product="thiamine biosynthesis lipoprotein" /protein_id="AAP95355.1" /db_xref="GI:33147834" /translation="MLVCVASLSFILTACNKSAEQITLOGKTMGTYTYVKYIDDGKIA NLPTAQVKQLDQVLKVNDEMSITQLDSQISRFNKREVNTPFEISPDFATVAAEA IRLNKITEGALDVTGFLVGPDRDLDPQSAAQINQRAAFVIGIDKLAINTBEHK PTUTKSPYNTLAKGFGVDKIADQLEMLGLENLYLEVGEURGKGNLQGLNW QIAIEKPNPEOGQTVITVPLNHLGMATSGNYRNPEDQGNRLSHIIDPKALKPWNH KLASITVLAPTSMTADGLSTGLFVLGAERKALEVAEREKLAIFLIINKGDPGPEQMSQQ FKLLIDQK"	
	gene	1155. .1403 /locus tag="HD0387" 1155. .1403 /locus tag="HD0387" /codon_start=1 /transl_table=11 /product="conserved hypothetical protein" /protein_id="AAP95356.1" /db_xref="GI:33147835" /translation="METLLTFFGFLVIFAMSIGFIIRKGIKSGCGGITALGMKKQM CDCEPCDNLKTKIMEGKADPAKIAQFKKXVASFYEVK"
gene	1155. .1403 /locus tag="HD0388" 1155. .1403 /locus tag="HD0388" /codon_start=1 /transl_table=11 /product="ton-dependent heme receptor A" /protein_id="AAP95357.1" /db_xref="GI:33147836"	
	CDS	1155. .1403 /locus tag="HD0389" 1155. .1403 /locus tag="HD0389" /gene="trmU"


```
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAP95362.1"
/db_xref="GI:33147841"
/translation="MNSIRFCSQIVQVPEPRTSLAFITGPGCKGCHSDSWKA
GSOILTESVLAELKYLRLISCVPMGEHLQKLLSLLKLVKQAGLHCLYGLLE
QSLPADLAELTYLKTGRWIAERGLTSLITNQHFIDLTGLENLHLFIKEFR"
10039..10620
/gene="rnfA"
/locus_tag="HD0396"
/gene="rnfA"
/locus_tag="HD0396"
/codon_start=1
/transl_table=11
/product="probable Na+-translocating NADH-quinone
oxidoreductase"
/protein_id="AAP95363.1"
/db_xref="GI:33147842"
/translation="WDYIILVISTALINNFVLVFLGLCPFMGVSKVETAIQMGA
TFVFLASLAVLIRFLLIPDLQFLRLTVFLVILVIVQTEMIVHKTSPTLRL
LGYELIITNCVAVGLNVLNLSNLSVLYGFGAAVGFSLVLVLFALRLAA
ADVRPFQASALITAGLMSLAFMGFTGLVKL"
10620..11243
/gene="rnfB"
/locus_tag="HD0397"
/gene="rnfB"
/locus_tag="HD0397"
/locus_tag="HD0397"
/gene="rnfB"
/locus_tag="HD0397"
/codon_start=1
/transl_table=11
/product="putative ferredoxin II, iron sulfur protein"
/protein_id="AAP95364.1"
/db_xref="GI:33147843"
/translation="MLDLSIIAYLLIALLIFGALLGYFSVKLVKVEADPIVDI
AIIPOSGCGGPGKPVAEATANGDOITKCPGPOPLVVKIAELMGLVDVPSMDNTA
VPEVKALLHEDMCIGCTKCIQACPVDAILIGNKAMHTVVDLCTGCELCVAPCPINC
IEMIKVQKPRSNWQFNPDILIIPIVNTTELQKQIMVGSAGEIQND"
11236..11344
/gene="rnfC"

Query Match 100.0%; Score 1168; DB 1; Length 301191;
Best Local Similarity 100.0%; Pred. No. 8.1e-206;
Matches 1168; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AATAATACGTCATTGACATTTTATGTAAGTAGAATAAGAAAGTAAATCTATATT 60
DB 298759 AATAATACGTCATTGACATTTTATGTAAGTAGAATAAGAAAGTAAATCTATATT 298818

QY 61 TACAATCAAGATTGACAAATTTATTTACTTAATAGAGGTGATTATGAAATTTAAATGTTTAGT 120
DB 298819 TACAATCAAGATTGACAAATTTATTTACTTAATAGAGGTGATTATGAAATTTAAATGTTTAGT 298878

QY 121 TCGCGTAGTGGGATTAGCTGTCTACTATTACAAGTGGCTCAGCGCGCAAGTT 180
DB 298879 TCGCGTAGTGGGATTAGCTGTCTACTATTACAAGTGGCTCAGCGCGCAAGTT 298938

QY 181 TCGTGGAGTATCTCTTTGTATAGCTATGATGATGACTATGGTAAGGTAAGTAAATGGACTTG 240
DB 298939 TCGTGGAGTATCTCTTTGTATAGCTATGATGATGACTATGGTAAGGTAAGTAAATGGACTTG 298998

QY 241 GTCTAATGAAGCGGTTTCGATATTAAGTGCCAGGATTAAATGAAGCCAAAGAATG 300
DB 298999 GTCTAATGAAGCGGTTTCGATATTAAGTGCCAGGATTAAATGAAGCCAAAGAATG 299058

QY 301 GATTCTTAACAGGCTACTTATCTTGAATTTACAGCATTTATGCTTATACCTCTGTTCT 360
DB 299059 GATTCTTAACAGGCTACTTATCTTGAATTTACAGCATTTATGCTTATACCTCTGTTCT 299118

QY 361 CGTGACATATGCTCCGCGCTTCTCCTAGCCCTTACTGTTATATCCGATGCTGATCC 420
DB 299119 CGTGACATATGCTCCGCGCTTCTCCTAGCCCTTACTGTTATATCCGATGCTGATCC 299178
```

```
QY 421 TGATCAACTTGGATAAATCGCAGCAGCTGAAATTTGAATTTGTATAGTTATTTTAAACGA 480
DB 299179 TGATCAACTTGGATAAATCGCAGCAGCTGAAATTTGAATTTGTATAGTTATTTTAAACGA 299238

QY 481 TTTAAGACACGATTTTAAATTTAAAGTCTTCTGATGCACGATTTTCCAAAATAAACAATA 540
DB 299239 TTTAAGACACGATTTTAAATTTAAAGTCTTCTGATGCACGATTTTCCAAAATAAACAATA 299298

QY 541 TATTGATCTAATAAGTAATATTACTAGAACTGGGTACTATTATTTAGATGATCTTATCG 600
DB 299299 TATTGATCTAATAAGTAATATTACTAGAACTGGGTACTATTATTTAGATGATCTTATCG 299358

QY 601 TATGATGGAACAAAATACACATAATCAATAAGTTGTCTAAAGAAATTCGAAACTGTTT 660
DB 299359 TATGATGGAACAAAATACACATAATCAATAAGTTGTCTAAAGAAATTCGAAACTGTTT 299418

QY 661 AGCCAACCAATCAGCAATCTCTATGTTAGTCAACCAAAATGGTGTAGGCAAAACGAGCGT 720
DB 299419 AGCCAACCAATCAGCAATCTCTATGTTAGTCAACCAAAATGGTGTAGGCAAAACGAGCGT 299478

QY 721 TTCTGCTCGGTAGAGGTTATAGAGATAAAACTGCAATAGCCATTTGGTGTGCGCTCAGC 780
DB 299479 TTCTGCTCGGTAGAGGTTATAGAGATAAAACTGCAATAGCCATTTGGTGTGCGCTCAGC 299538

QY 781 CATTACTGATCGCTTTTACCCTAAAGCGGTGTAGCGTTCAATACCTCAATGCGCGCAT 840
DB 299539 CATTACTGATCGCTTTTACCCTAAAGCGGTGTAGCGTTCAATACCTCAATGCGCGCAT 299598

QY 841 GTCCTATGCTGCTTCTGTTGGTTATGAATTTCAATCATTACGTTTAAATCACTAATCGTTT 900
DB 299599 GTCCTATGCTGCTTCTGTTGGTTATGAATTTCAATCATTACGTTTAAATCACTAATCGTTT 299658

QY 901 TGGTTATAATAAAAGCGTAAATGTTTCTCCTCACATTTAGCCTTTCTTATTTATCTTTG 960
DB 299659 TGGTTATAATAAAAGCGTAAATGTTTCTCCTCACATTTAGCCTTTCTTATTTATCTTTG 299718

QY 961 TTATAGCTTTTCTGTTTATAAAACCGTTTGTAGCCACATTTTATTAATTAAGCTTTTAAAG 1020
DB 299719 TTATAGCTTTTCTGTTTATAAAACCGTTTGTAGCCACATTTTATTAATTAAGCTTTTAAAG 299778

QY 1021 CCTATTCAATCAGTCTTACTTTTACCTTTTCCACCATTTATTCGCCACTTCTTAAACGG 1080
DB 299779 CCTATTCAATCAGTCTTACTTTTACCTTTTCCACCATTTATTCGCCACTTCTTAAACGG 299838

QY 1081 TAATAATTAAGTTGGTTTACCTTAAATGGGTACCTTCTATCGGAAATTTTCTTAAATGTT 1140
DB 299839 TAATAATTAAGTTGGTTTACCTTAAATGGGTACCTTCTATCGGAAATTTTCTTAAATGTT 299898

QY 1141 CTAAATTTAAGCGTTTAAAGTGGCGAC 1168
DB 299899 CTAAATTTAAGCGTTTAAAGTGGCGAC 299926
```

RESULT 2

AF187001

LOCUS

DEFINITION

Haemophilus ducreyi strain 35000 serum resistance protein Dsra

(dsra) gene, complete cds.

ACCESSION

AF187001

VERSION

AF187001.1

GI:7188572

KEYWORDS

SOURCE

ORGANISM

Haemophilus ducreyi

Haemophilus ducreyi

Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;

Pasteurellaceae; Haemophilus.

1 (bases 1 to 1000)

Elkins, C., Morrow, K.J. Jr. and Olsen, B.

Serum resistance in Haemophilus ducreyi requires outer membrane

protein Dsra

Infect. Immun. 68 (3), 1608-1619 (2000)

JOURNAL

MEDLINE

PUBMED

10678980

```
REFERENCE 2 (bases 1 to 1000)
AUTHORS Elkins,C., Morrow,K.J. and Olsen,B.
TITLE Direct Submission
JOURNAL Submitted (17-SEP-1999) Division of Infectious Diseases, Department
of Medicine, University of North Carolina at Chapel Hill, 521
Burnett-Womack Bldg. CB7030, Chapel Hill, NC 27516, USA

FEATURES
source
1..1000
/organism="Haemophilus ducreyi"
/mol_type="genomic DNA"
/strain="35000"
/db_xref="taxon:730"
/notes="serum resistant strain"
101..874
/gene="dsrA"
101..874
/gene="dsrA"
/notes="outer membrane protein required for serum
resistance"
/codon_start=1
/transl_table=11
/product="serum resistance protein DsrA"
/protein_id="AAF37807.1"
/db_xref="GI:7188573"
/translation="MKIKCLVAVVGLACSTTTMAOQPPKPGVSSLSYEDYGGK
WTNSGEGFDIKVPIKMKPKEMISKQATYLELQHYMPYTVLVTYAPGVSPSPILLY
PMSDDOLGINROQLKLNLYSPFNDLRHDKLVLDARISKKNIDTIISKYLLLEGT
YLDOSYRMFQNTNHNKLSKELDTGLANOSALSMVQPNVGKTSVSAVGGYRDKT
ALAGVGSRIIDRTFAKAGVAFNYNGMSYGSVGYEF"

ORIGIN
Query Match 85.6%; Score 1000; DB 1; Length 1000;
Best Local Similarity 100.0%; Pred. No. 2.3e-174;
Matches 1000; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATAAATACGTCATTGACATTTTAAATGTAAGGTAGAATAAGAAAGTAATTTCTATTT 60
Db 1 ATAAATACGTCATTGACATTTTAAATGTAAGGTAGAATAAGAAAGTAATTTCTATTT 60

Qy 61 TACAATCAAGATTGCACAAATTTTACTTAATGAGGTGATATGAAATTAATCTTTAGT 120
Db 61 TACAATCAAGATTGCACAAATTTTACTTAATGAGGTGATATGAAATTAATCTTTAGT 120

Qy 121 TGCCTAGTGGGATTAGCTTCTTCTACTATTACAAATGGCTCAGCAGCGCCAAAGTT 180
Db 121 TGCCTAGTGGGATTAGCTTCTTCTACTATTACAAATGGCTCAGCAGCGCCAAAGTT 180

Qy 181 TGCTGGAGTATCTCTTTGTATAGCTATGAGTATGACTATGGTAAAGGTAATGGACTTG 240
Db 181 TGCTGGAGTATCTCTTTGTATAGCTATGAGTATGACTATGGTAAAGGTAATGGACTTG 240

Qy 241 GTCTAATGAAGCGGTTTCGATATTAAAGTCCAGGGAATTAATGAAGCCAAAGATG 300
Db 241 GTCTAATGAAGCGGTTTCGATATTAAAGTCCAGGGAATTAATGAAGCCAAAGATG 300

Qy 301 GATTTCCTAAACAGGCTACTTATCTTTGAATTAACAGCATTATATGCTTATATCTCTTCT 360
Db 301 GATTTCCTAAACAGGCTACTTATCTTTGAATTAACAGCATTATATGCTTATATCTCTTCT 360

Qy 361 CGTGACATATGCTCTCGCGTTTCTCTAGCCCTATCTGTTATATCCGATGTCGTATCC 420
Db 361 CGTGACATATGCTCTCGCGTTTCTCTAGCCCTATCTGTTATATCCGATGTCGTATCC 420

Qy 421 TGATCACTTGGATAAATCGCAGCAGCTCAAAATGAAATTTGATATGTTTATTAACGA 480
Db 421 TGATCACTTGGATAAATCGCAGCAGCTCAAAATTTGATATGTTTATTAACGA 480

Qy 481 TTTAAGACACGATTTTAAATTAAGTTCTTTGATGACGATTTTCCAAAATAAACAAA 540
Db 481 TTTAAGACACGATTTTAAATTAAGTTCTTTGATGACGATTTTCCAAAATAAACAAA 540

Qy 541 TATTGATCTATAGTAAATTTTACTAGAACTGGGTACTTATTTAGATGATTTCTTATCG 600
Db 541 TATTGATCTATAGTAAATTTTACTAGAACTGGGTACTTATTTAGATGATTTCTTATCG 600
```

```
Db 541 TATTGATCTATAGTAAATTTTACTAGAACTGGGTACTTATTTAGATGATTTCTTATCG 600
Qy 601 TATGATGGACAAAATACACATATATCAATAAGTTGTCTTAAAGAAATGCAAACTGGTTT 660
Db 601 TATGATGGACAAAATACACATATATCAATAAGTTGTCTTAAAGAAATGCAAACTGGTTT 660
Qy 661 AGCAACCAATCAGCATTTGTCTATGTTAGTGCAACCAATGGTGTAGGCAAAACGAGCGT 720
Db 661 AGCAACCAATCAGCATTTGTCTATGTTAGTGCAACCAATGGTGTAGGCAAAACGAGCGT 720
Qy 721 TTCTGTCGGGTAGGAGTTATAGAGATAAATGCAATAGGCATTTGGTTCGGCTCAGC 780
Db 721 TTCTGTCGGGTAGGAGTTATAGAGATAAATGCAATAGGCATTTGGTTCGGCTCAGC 780
Qy 781 CATTACTGATCGCTTTACCGCTAAAGCGGCTGAGCGTTCAATACCTACCAATGGCGGCAT 840
Db 781 CATTACTGATCGCTTTACCGCTAAAGCGGCTGAGCGTTCAATACCTACCAATGGCGGCAT 840
Qy 841 GTCTTATGGTCTTCTGTTTATGAATTTCAATACATTACGTTTAACTAATCGTTT 900
Db 841 GTCTTATGGTCTTCTGTTTATGAATTTCAATACATTACGTTTAACTAATCGTTT 900
Qy 901 TGGTTATAATAAAGCTTAAATGTTTCTCCTCACATTTAGCCCTTCTTATTTATCTTTG 960
Db 901 TGGTTATAATAAAGCTTAAATGTTTCTCCTCACATTTAGCCCTTCTTATTTATCTTTG 960
Qy 961 TTATAGCTTTTGTGTTTATATAAACCGTTTTTTTAGCCACTT 1000
Db 961 TTATAGCTTTTGTGTTTATATAAACCGTTTTTTTAGCCACTT 1000

RESULT 3
AF187005 874 bp DNA linear BCT 07-MAR-2000
LOCUS Haemophilus ducreyi strain 406 serum resistance protein DsrA (dsrA)
DEFINITION gene, complete cds.
ACCESSION AF187005
VERSION AF187005.1 GI:7188580
KEYWORDS Haemophilus ducreyi
SOURCE Haemophilus ducreyi
ORGANISM Haemophilus ducreyi
Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
Pasteurellaceae; Haemophilus.
REFERENCE 1 (bases 1 to 874)
AUTHORS Elkins,C., Morrow,K.J. Jr. and Olsen,B.
TITLE Serum resistance in Haemophilus ducreyi requires outer membrane
protein DsrA
JOURNAL Infect. Immun. 68 (3), 1608-1619 (2000)
MEDLINE 20143779
PUBMED 10678980
REFERENCE 2 (bases 1 to 874)
AUTHORS Elkins,C., Morrow,K.J. and Olsen,B.
TITLE Direct Submission
JOURNAL Submitted (17-SEP-1999) Division of Infectious Diseases, Department
of Medicine, University of North Carolina at Chapel Hill, 521
Burnett-Womack Bldg. CB7030, Chapel Hill, NC 27516, USA

FEATURES
source
1..874
/organism="Haemophilus ducreyi"
/mol_type="genomic DNA"
/strain="406"
/db_xref="taxon:730"
/notes="serum resistant strain"
101..874
/gene="dsrA"
101..874
/gene="dsrA"
/notes="outer membrane protein required for serum
resistance"
/codon_start=1
/transl_table=11
/product="serum resistance protein DsrA"
/protein_id="AAF37811.1"
```


Db 241 GTCTAATGAAGCGGTTTCGATATATAAGTGCCAGGGATTAATAAGTGAAGCCAAAGAATG 300
Qy 301 GATTCTTAAACAGCGCTACTTATCTTGAATACAGCATTTATATGCTTATATCTCTGTTCT 360
Db 301 GATTCTTAAACAGCGCTACTTATCTTGAATACAGCATTTATATGCTTATATCTCTGTTCT 360
Qy 361 CGTGACATATGCTCCTGCGGTTCTCCTAGCCCTATATCTGTATATCCGATGTCGATCC 420
Db 361 CGTGACATATGCTCCTGCGGTTCTCCTAGCCCTATATCTGTATATCCGATGTCGATCC 420
Qy 421 TGATCAACTTGGATAAATCGGAGCAGCTGAATTTGAATTTGTATATTTTAAACGA 480
Db 421 TGATCAACTTGGATAAATCGGAGCAGCTGAATTTGTATATTTTAAACGA 480
Qy 481 TTTAAGCACACATTTTAAATTTAAAGTCTTGATGACGATTTTCCAAAAATAACAAA 540
Db 481 TTTAAGCACACATTTTAAATTTAAAGTCTTGATGACGATTTTCCAAAAATAACAAA 540
Qy 541 TATTGATACTATAAGTAAATTTTACTAGAACTGGGTACTTATTTAGATGATTTCTATCG 600
Db 541 TATTGATACTATAAGTAAATTTTACTAGAACTGGGTACTTATTTAGATGATTTCTATCG 600
Qy 601 TATGATGAACAAATACACATATATCAATTAAGTTGTCTAAAGAAATGCAAACTGGTTT 660
Db 601 TATGATGAACAAATACACATATATCAATTAAGTTGTCTAAAGAAATGCAAACTGGTTT 660
Qy 661 AGCCAAACCAATCAGCATTTCTATGTTAGTGCACCAAAATGGTGTAGGCAAAACGAGCGT 720
Db 661 AGCCAAACCAATCAGCATTTCTATGTTAGTGCACCAAAATGGTGTAGGCAAAACGAGCGT 720
Qy 721 TTCTGCTGCGTAGGAGGTTATAGAGATPAAATCGCATTTAGCCATTTGGTGTGCGCTCACG 780
Db 721 TTCTGCTGCGTAGGAGGTTATAGAGATPAAATCGCATTTAGCCATTTGGTGTGCGCTCACG 780
Qy 781 CATTAAGTATCGCTTTACCGCTAAAGCGGTTGTAGCGTTTCAATACCTACAAATGCGGCGAT 840
Db 781 CATTAAGTATCGCTTTACCGCTAAAGCGGTTGTAGCGTTTCAATACCTACAAATGCGGCGAT 840
Qy 841 GTCTTATGCTCTCTGTTGGTTATGAATTTCTAA 874
Db 841 GTCTTATGCTCTCTGTTGGTTATGAATTTCTAA 874

RESULT 5
AF187004
LOCUS AF187004 847 bp DNA linear BCT 07-MAR-2000
DEFINITION Haemophilus ducreyi strain CIP542(Can) serum resistance protein
DsrA (dsrA) gene, complete cds.
ACCESSION AF187004
VERSION AF187004.1 GI:7188578
KEYWORDS
SOURCE Haemophilus ducreyi
ORGANISM Haemophilus ducreyi
Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
Pasteurellaceae; Haemophilus.
REFERENCE 1 (bases 1 to 847)
AUTHORS Elkins, C., Morrow, K. J., Jr. and Olsen, B.
TITLE Serum resistance in Haemophilus ducreyi requires outer membrane protein DsrA
JOURNAL Infect. Immun. 68 (3), 1608-1619 (2000)
MEDLINE 20143779
PUBMED 10678980
REFERENCE 2 (bases 1 to 847)
AUTHORS Elkins, C., Morrow, K. J. and Olsen, B.
TITLE Direct Submission
JOURNAL Submitted (17-SEP-1999) Division of Infectious Diseases, Department of Medicine, University of North Carolina at Chapel Hill, 521 Burnett-Womack Bldg. CB7030, Chapel Hill, NC 27516, USA
FEATURES
source
1...847
/organism="Haemophilus ducreyi"
/mol_type="genomic DNA"
/strain="CIP542(Can)"

/db_xref="taxon:730"
/notes="serum susceptible strain; obtained from William Albritton"
101...847
/genes="dsrA"
101...847
/genes="dsrA"
/notes="outer membrane protein required for serum resistance; deletion within coding region results in premature stop codon"
/codon_start=1
/transl_table=11
/product="serum resistance protein DsrA"
/protein_id="AAF37810.1"
/db_xref="GI:7188579"
/translations="MKIKLVAVVGLACSTTTMAQPPKPFAGVSSILSYEYDYGKKG
WTWSNGGDDIKVPKINPKKWKISQATYLEQHYMPYTPVLVYAPGVSPPILLY
PMSDDPOLGINROQLNLYSYFNDFDKLRLDKARISKNQNDITISKYLELGT
YLDSSYRMMEQNTHNINKSKELOQTGLANQALSMLVQPNVGKTSVSAVGGYRDKT
ALAIGVGSRIITDRFTAKAGVAFNFTYRNF"
ORIGIN
Query Match 71.0%; Score 829.2; DB 1; Length 847;
Best Local Similarity 99.6%; Pred. No. 6.8e-143;
Matches 831; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Qy 1 ATAAATACGTCATTGACATTTTATTAATGTAAGTAGAATAAGAAAGTAAATTTCTATTT 60
Db 1 ATAAATACGTCATTGACATTTTATTAATGTAAGTAGAATAAGAAAGTAAATTTCTATTT 60
Qy 61 TACAATCAAGATTGACAAATTTATTTACTTAATGAGGTGATTTGAAAAATTAATTTAGT 120
Db 61 TACAATCAAGATTGACAAATTTATTTACTTAATGAGGTGATTTGAAAAATTAATTTAGT 120
Qy 121 TGCGTAGTGGGATTAGCTTTGTTCTACTATTACAACATGCTCAGCAGCGCCAAAGATT 180
Db 121 TGCGTAGTGGGATTAGCTTTGTTCTACTATTACAACATGCTCAGCAGCGCCAAAGATT 180
Qy 181 TGCTGGAGTATCTTTGTTGATAGCTATGACTATGCTTAAAGGTAAATCGACTTG 240
Db 181 TGCTGGAGTATCTTTGTTGATAGCTATGACTATGCTTAAAGGTAAATCGACTTG 240
Qy 241 GTCTAATGAAGCGGTTTCGATATTTAAAGTCCAGGATTTAAATGAAGCCAAAGAATG 300
Db 241 GTCTAATGAAGCGGTTTCGATATTTAAAGTCCAGGATTTAAATGAAGCCAAAGAATG 300
Qy 301 GATTCTTAAACAGCGCTACTTATCTTGAATACAGCATTTATATGCTTATATCTCTGTTCT 360
Db 301 GATTCTTAAACAGCGCTACTTATCTTGAATACAGCATTTATATGCTTATATCTCTGTTCT 360
Qy 361 CGTGACATATGCTCCTGCGGTTTCTCCTAGCCCTATATCTGTATATCCGATGTCGATCC 420
Db 361 CGTGACATATGCTCCTGCGGTTTCTCCTAGCCCTATATCTGTATATCCGATGTCGATCC 420
Qy 421 TGATCAACTTGGATAAATCGGAGCAGCTGAAATTTGAATTTGTATATTTTAAACGA 480
Db 421 TGATCAACTTGGATAAATCGGAGCAGCTGAAATTTGTATATTTTAAACGA 480
Qy 481 TTTAAGCACAGATTTTAAATTTAAAGTCTTGATGACGATTTTCCAAAAATAACAAA 540
Db 481 TTTAAGCACAGATTTTAAATTTAAAGTCTTGATGACGATTTTCCAAAAATAACAAA 540
Qy 541 TATTGATACTATAAGTAAATTTTACTAGAACTGGGTACTTATTTAGATGATTTCTATCG 600
Db 541 TATTGATACTATAAGTAAATTTTACTAGAACTGGGTACTTATTTAGATGATTTCTATCG 600
Qy 601 TATGATGAACAAATACACATATATCAATTAAGTTGTCTAAAGAAATGCAAACTGGTTT 660
Db 601 TATGATGAACAAATACACATATATCAATTAAGTTGTCTAAAGAAATGCAAACTGGTTT 660
Qy 661 AGCCAAACCAATCAGCATTTGTCTATGTTAGTGCACCAAAATGGTGTAGGCAAAACGAGCGT 720
Db 661 AGCCAAACCAATCAGCATTTGTCTATGTTAGTGCACCAAAATGGTGTAGGCAAAACGAGCGT 720

gene
CDS

```

Qy 721 TTCTGTCGGGTAGGAGGTTATAGAGATAAACTGCAATTAGCCATTGGTGTGGCTCACG 780
Db 721 TTCTGTCGGGTAGGAGGTTATAGAGATAAACTGCAATTAGCCATTGGTGTGGCTCACG 780
Qy 781 CATTACTGATCGCTTTTACCCTGTAAGCGGGGTGTAGCCCTCAATACCTACCAATGG 834
Db 781 CATTACTGATCGCTTTTACCCTGTAAGCGGGGTGTAGCCCTCAATACCTTCTATCG 834

RESULT 6
AF187002 890 bp DNA linear BCT 07-MAR-2000
LOCUS Haemophilus ducreyi strain CIP475 serum resistance protein DsrA
DEFINITION (dsrA) gene, complete cds.
ACCESSION AF187002
VERSION AF187002.1 GI:7188574
KEYWORDS
SOURCE Haemophilus ducreyi
ORGANISM Haemophilus ducreyi
Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
Pasteurellaceae; Haemophilus.
REFERENCE 1 (bases 1 to 890)
AUTHORS Elkins, C., Morrow, K.J. Jr. and Olsen, B.
TITLE Serum resistance in Haemophilus ducreyi requires outer membrane
protein DsrA
JOURNAL Infect. Immun. 68 (3), 1608-1619 (2000)
MEDLINE 20143779
PUBMED 10678980
REFERENCE 2 (bases 1 to 890)
AUTHORS Elkins, C., Morrow, K.J. and Olsen, B.
TITLE Direct Submission
JOURNAL Submitted (17-SEP-1999) Division of Infectious Diseases, Department
of Medicine, University of North Carolina at Chapel Hill, 521
Burnett-Womack Bldg. CB7030, Chapel Hill, NC 27516, USA
FEATURES
source
1..890
/organism="Haemophilus ducreyi"
/mol_type="genomic DNA"
/strain="CIP475"
/db_xref="taxon:730"
/notes="serum susceptible strain"
96..890
/gene="dsrA"
96..890
/gene="dsrA"
/notes="outer membrane protein required for serum
resistance; no expression of DsrA in this strain due to a
5 bp deletion between putative -10 and -35 promoter
regions"
/codon_start=1
/transl_table=11
/product="serum resistance protein DsrA"
/protein_id="AAF37808.1"
/db_xref="GI:7188575"
/translations="MKIKLVAVVGLACSTITWAOQPKFAGVSLSYEVDYKQK
WTWSNEGDFIKVPQIKMKPEWISKQATLEQHYNPYPPLVLTVAHDVPPSSILLY
PMSDQPLQGLKNLKYSYFNDLRHDFKVLDRISKQNDITSKYLLELGT
YLDSSYRMEEQNTNINKNTNINKLSKELOTLGANSLSMLVQPNVGKTSVSAV
GGYRDKTALAIGVSRITDRFTAKAGVAFNTYNGMSYGSVGYEF"

Query Match 69.9%; Score 816.6; DB 1; Length 890;
Best Local Similarity 96.6%; Pred. No. 1.4e-140;
Matches 865; Conservative 0; Mismatches 4; Indels 26; Gaps 2;

Qy 1 ATAAATACGTCATTGACATTTTATGTAAGGTAGCAATAAGAAAGTAATCTATATT 60
Db 1 ATAAATACGTCATTGACATTTT-----TAAGGTAGCAATAAGAAAGTAATCTATATT 55
Qy 61 TACAATCAAGATTGACAAATTATTACTTAATGAGGTGATTATGAATAATTAATGTTTGTAGT 120
Db 56 TACAATCAAGATTGACAAATTATTACTTAATGAGGTGATTATGAATAATTAATGTTTGTAGT 115

```

```

Qy 121 TGCCGTAGTGGGANTTAGCTTGTCTACTATTACAACTGGCTCAGCAGCGCCGCAAGTT 180
Db 116 TGCCGTAGTGGGANTTAGCTTGTCTACTATTACAACTGGCTCAGCAGCGCCGCAAGTT 175
Qy 181 TGCTGGAGTATCTTCTTGTATAGCTATGATGATGACTATGGTAAAGGTAATGGACTTG 240
Db 176 TGCTGGAGTATCTTCTTGTATAGCTATGATGATGACTATGGTAAAGGTAATGGACTTG 235
Qy 241 GTCTAATGAAGCGGTTTCGATATTAAAGTCCAGGATTAAATGAAGCAAAAGATG 300
Db 236 GTCTAATGAAGCGGTTTCGATATTAAAGTCCAGGATTAAATGAAGCAAAAGATG 295
Qy 301 GATTTCTTAACAGCGCTACTTATCTTGAATTTACAGCATTTATGCCCTTATCTCTGTCT 360
Db 296 GATTTCTTAACAGCGCTACTTATCTTGAATTTACAGCATTTATGCCCTTATCTCTGTCT 355
Qy 361 CGTGACATATGCTTCCTGGGCTTCTCTAGCCCTTATCTAGCTTATATCCGATGCTGATCC 420
Db 356 CGTGACATATGCTTCATGACGTTCTCTCTAGCTTATATCCGATGCTGATCC 415
Qy 421 TGATCAACTTTGGAATAAATCGGCAGCAGCTGAAATTTGAATTTGTATAGTATTATTAACGA 480
Db 416 TGATCAACTTTGGAATAAATCGGCAGCAGCTGAAATTTGAATTTGTATAGTATTATTAACGA 475
Qy 481 TTTAAGACACGATTTTAAATTTAAAGTCTTGTGATGACGCTATTTCCAAAATAAACAATA 540
Db 476 TTTAAGACACGATTTTAAATTTAAAGTCTTGTGATGACGCTATTTCCAAAATAAACAATA 535
Qy 541 TATTGATCTATATAAGTAAATATTATTAGAACTGGGTACTTTATTAGATGATCTTATATCG 600
Db 536 TATTGATCTATATAAGTAAATATTATTAGAACTGGGTACTTTATTAGATGATCTTATATCG 595
Qy 601 TATGATGGAAAC-----AAAATACACATAATATCAATAAATAATACATAATATCAATAAGTTGTC 639
Db 596 TATGATGGAAACAAAATACACATAATATCAATAAATAATACATAATATCAATAAGTTGTC 655
Qy 640 TAAAGAAATTGCAAACTGGTTTAGCCCAACCAATCAGCATTTGCTATCTTGTAGTGAACAAA 699
Db 656 TAAAGAAATTGCAAACTGGTTTAGCCCAACCAATCAGCATTTGCTATGTTAGTGAACAAA 715
Qy 700 TGGGTAGGCAAAACGAGCGTTTCTGCTCGGTAGGAGTTATAGAGATAAATACTGCATT 759
Db 716 TGGGTAGGCAAAACGAGCGTTTCTGCTCGGTAGGAGTTATAGAGATAAATACTGCATT 775
Qy 760 AGCCATTGGTGTGGCTCAGCATTTACTGATCGCTTTTACCGCTAAAGCGGGTGTAGCGTT 819
Db 776 AGCCATTGGTGTGGCTCAGCATTTACTGATCGCTTTTACCGCTAAAGCGGGTGTAGCGTT 835
Qy 820 CAATACCTACAATGGCGCATGCTTATGTTGCTTCTGTTGTTATGAATTTCTAA 874
Db 836 CAATACCTACAATGGCGCATGCTTATGTTGCTTCTGTTGTTATGAATTTCTAA 890

RESULT 7
AF187003 890 bp DNA linear BCT 07-MAR-2000
LOCUS Haemophilus ducreyi strain CIP477 serum resistance protein DsrA
DEFINITION (dsrA) gene, complete cds.
ACCESSION AF187003
VERSION AF187003.1 GI:7188576
KEYWORDS
SOURCE Haemophilus ducreyi
ORGANISM Haemophilus ducreyi
Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
Pasteurellaceae; Haemophilus.
REFERENCE 1 (bases 1 to 890)
AUTHORS Elkins, C., Morrow, K.J. Jr. and Olsen, B.
TITLE Serum resistance in Haemophilus ducreyi requires outer membrane
protein DsrA
JOURNAL Infect. Immun. 68 (3), 1608-1619 (2000)
MEDLINE 20143779
PUBMED 10678980

```

REFERENCE 2 (bases 1 to 890)
AUTHORS Elkins, C., Morrow, K.J. and Olsen, B.
TITLE Direct Submission
JOURNAL Submitted (17-SEP-1999) Division of Infectious Diseases, Department of Medicine, University of North Carolina at Chapel Hill, 521 Burnett-Womack Bldg. CB7030, Chapel Hill, NC 27516, USA

FEATURES
source
1..890
/organism="Haemophilus ducreyi"
/mol_type="genomic DNA"
/strain="CIPA77"
/db_xref="taxon:730"
/note="serum susceptible strain"
96..890
/gene="dsrA"
96..890
/gene="dsrA"
/note="outer membrane protein required for serum resistance; little or no expression of DsrA in this strain"
/codon_start=1
/transl_table=11
/product="serum resistance protein DsrA"
/protein_id="AAF37809.1"
/db_xref="GI:7188577"

WTNSNGGFDIKVPGIKMKPKWISKOATYLELQHYNPYLPVLVTSAPDVPSSILLY
PMSDDQLGNRQOLKLNLYSYFNDLRHDFKLVDARISKNKNIDITISKILLELGT
YLDGYSRMEQNTHNKNTHNKNLSKELOTLGANSLSMLVQPNVGKTSVSAV
GGYRDKTALAIGVSRITDRFTAKAGVAFNTYNGMSYGASGVYEF"

ORIGIN
Query Match 69.9%; Score 816.6; DB 1; Length 890;
Best Local Similarity 96.6%; Pred. No. 1.4e-140;
Matches 865; Conservative 0; Mismatches 4; Indels 26; Gaps 2;
QY 1 ATAAATACGTCATTGACATTTTATTAAGTAAAGTAGAATAAGAAAGTAAATTCATATT 60
DB 1 ATAAATACGTCATTGACATTTTATTAAGTAAAGTAGAATAAGAAAGTAAATTCATATT 55
QY 61 TACAATCAAGATTGACAAATTTTACTTAATGAGGTGATTATGAAAAATTAATGTTTAGT 120
DB 56 TACAATCAAGATTGACAAATTTTACTTAATGAGGTGATTATGAAAAATTAATGTTTAGT 115
QY 121 TGGCGTAGTGGGATTAGCTTCTTACTATTACAAATGCTCAGACGCCGCAAGATT 180
DB 116 TGGCGTAGTGGGATTAGCTTCTTACTATTACAAATGCTCAGACGCCGCAAGATT 175
QY 181 TGCTGGAGTATCTTCTTTGTATAGCTATGAGTATGACTATGCTAAGGTTAAATGCACTTG 240
DB 176 TGCTGGAGTATCTTCTTTGTATAGCTATGAGTATGACTATGCTAAGGTTAAATGCACTTG 235
QY 241 GTCTAATGAAGCGCGTTTCGATATTAAAGTCCAGGATTAATAATGAAGCAAAAGAAATG 300
DB 236 GTCTAATGAAGCGCGTTTCGATATTAAAGTCCAGGATTAATAATGAAGCAAAAGAAATG 295
QY 301 GATTCTTAAACAGGCTACTTATCTTGAATTAACAGATTAATGCTTATACCTCTGTCT 360
DB 296 GATTCTTAAACAGGCTACTTATCTTGAATTAACAGATTAATGCTTATACCTCTGTCT 355
QY 361 CGTGACATATCTCTCGCGTTTCTCCTPAGCCCTATCTGTTATATCCGATGCTGATCC 420
DB 356 CGTGACATATCTCTATGACGTTCTCTCTPAGCTCTATCTGTTATATCCGATGCTGATCC 415
QY 421 TGATCAACTTGGAAATAATCGGCAGCAGCTGAAATTTGAATTTGATGTTATTTTAAACGA 480
DB 416 TGATCAACTTGGAAATAATCGGCAGCAGCTGAAATTTGAATTTGATGTTATTTTAAACGA 475
QY 481 TTTAAGACACAGATTTTAAATTTAAAGTCTTGATGACGATATTTCCAAAAATAACAAA 540
DB 476 TTTAAGACACAGATTTTAAATTTAAAGTCTTGATGACGATATTTCCAAAAATAACAAA 535
QY 541 TATTGATACTATAAGTAAATTTTACTAGAACTGGGTACTTATTATTAGATGATTTCTTATCG 600

DB 536 TATTGATACTATAAGTAAATATTACTAGAACTGGGTACTTATTAGATGATTTCTATCG 595
QY 601 TATGATGGAACT-----AAATACACATAATATCAATAAGTTGTC 639
DB 596 TATGATGGAACT-----AAATACACATAATATCAATAAATATCAATAAGTTGTC 655
QY 640 TAAAGAAATGCAAACTGGTTTAGCCCAACCAATCAGCATTCCTATGTTAGTGCACCAAA 699
DB 656 TAAAGAAATGCAAACTGGTTTAGCCCAACCAATCAGCATTCCTATGTTAGTGCACCAAA 715
QY 700 TGGTGTAGGCAAAACGAGCGTTTCTGCTGCGGTAGGAGTTATAGAGATAAACTGCATT 759
DB 716 TGGTGTAGGCAAAACGAGCGTTTCTGCTGCGGTAGGAGTTATAGAGATAAACTGCATT 775
QY 760 AGCATTGCTGCTGCGCTCAGCATTTACTGCTTTTACCGCTAAACGGGTAGCGTT 819
DB 776 AGCATTGCTGCTGCGCTCAGCATTTACTGCTTTTACCGCTAAACGGGTAGCGTT 835
QY 820 CAATACCTACAATGGCGCATGCTTATGCTGCTTCTGTTGGTTATCAATTTCTAA 874
DB 836 CAATACCTACAATGGCGCATGCTTATGCTGCTTCTGTTGGTTATCAATTTCTAA 890

RESULT 8
AF187009
LOCUS Haemophilus ducreyi strain Vii57 serum resistance protein DsrA
DEFINITION (dsrA) gene, complete cds.
ACCESSION AF187009
VERSION AF187009.1 GI:7188588
KEYWORDS Haemophilus ducreyi
SOURCE Haemophilus ducreyi
ORGANISM Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales; Pasteurellaceae; Haemophilus.
REFERENCE 1 (bases 1 to 916)
AUTHORS Elkins, C., Morrow, K.J. Jr. and Olsen, B.
TITLE Serum resistance in Haemophilus ducreyi requires outer membrane protein DsrA
JOURNAL Infect. Immun. 68 (3), 1608-1619 (2000)
MEDLINE 20143779
PUBMED 10678980
REFERENCE 2 (bases 1 to 916)
AUTHORS Elkins, C., Morrow, K.J. and Olsen, B.
TITLE Direct Submission
JOURNAL Submitted (17-SEP-1999) Division of Infectious Diseases, Department of Medicine, University of North Carolina at Chapel Hill, 521 Burnett-Womack Bldg. CB7030, Chapel Hill, NC 27516, USA

FEATURES
source
1..916
/organism="Haemophilus ducreyi"
/mol_type="genomic DNA"
/strain="Vii57"
/db_xref="taxon:730"
/note="serum resistant strain"
101..916
/gene="dsrA"
101..916
/gene="dsrA"
/note="outer membrane protein required for serum resistance"
/transl_table=11
/product="serum resistance protein DsrA"
/protein_id="AAF37815.1"
/db_xref="GI:7188589"

WTNSNGGFDIKVPGIKMKPKWISKOATYLELQHYNPYLPVLVTSAPDVPSSILLY
PMSDDQLGNRQOLKLNLYSYFNDLRHDFKLVDARISKNKNIDITISKILLELGT
YLDGYSRMEQNTHNKNTHNKNLSKELOTLGANSLSMLVQPNVGKTSVSAV
GGYRDKTALAIGVSRITDRFTAKAGVAFNTYNGMSYGASGVYEF"

ORIGIN

Query Match	69.7%; Score 814; DB 1; Length 916;	
Best Local Similarity	94.9%; Pred. No. 4.2e-140;	
Matches	869; Conservative 0; Mismatches 5; Indels 42; Gaps 1;	
QY	1	ATAAATACGTCATGTGACATTTTATTAATGTAAGCTAGATAAGAAAGTAAATCTATATT 60
DB	1	ATAAATACGTCATGTGACATTTTATTAATGTAAGCTAGATAAGAAAGTAAATCTATATT 60
QY	61	TACAATCAAGATTGACAAATTTATTTACTTAATAGAGTGATTAATGAAATTAATGTTAGT 120
DB	61	TACAATCAAGATTGACAAATTTATTTACTTAATAGAGTGATTAATGAAATTAATGTTAGT 120
QY	121	TGCGTAGTGGGATTAGCTGTTCTACTATTACAAGTGGCTCAGCAGCGCCAAAGTT 180
DB	121	TGCGTAGTGGGATTAGCTGTTCTACTATTACAAGTGGCTCAGCAGCGCCAAAGTT 180
QY	181	TGCTGGAGTATCTTCTTTGATAGCTATGATGATGACTATGTAAGGTAATGGAAGTTG 240
DB	181	TGCTGGAGTATCTTCTTTGATAGCTATGATGATGACTATGTAAGGTAATGGAAGTTG 240
QY	241	GTCCTAATGAAGCGGTTTCGATATTAAGTGCCAGGATTAATGAAGCCAAAGAAATG 300
DB	241	GTCCTAATGAAGCGGTTTCGATATTAAGTGCCAGGATTAATGAAGCCAAAGAAATG 300
QY	301	GATTTCTAAACAGGCTACTTATCTTGAATTTACAGCATTAATGCTTATCTCTGTTCT 360
DB	301	GATTTCTAAACAGGCTACTTATCTTGAATTTACAGCATTAATGCTTATCTCTGTTCT 360
QY	361	CGTGACATATGCTCCTGGGTTCTCTAGCCCTATCTATCTTATCCGATGCTGATCC 420
DB	361	CGTGACATCTGCTCCTGACGTTCTCTAGCTCTATCTATCTTATCCGATGCTGATCC 420
QY	421	TGATCAACTTGGAAATAAATCGCAGCAGCTGAAATTTGATGATTTATTTAAACGA 480
DB	421	TGATCAACTTGGAAATAAATCGCAGCAGCTGAAATTTGATGATTTATTTAAACGA 480
QY	481	TTTAAGACACGATTTTAAATTAAGTTCTTGATGACGATTTTCGAAATTAACAAATA 540
DB	481	TTTAAGACACGATTTTAAATTAAGTTCTTGATGACGATTTTCGAAATTAACAAATA 540
QY	541	TATTGATCTATAGTAATATTTTACTAGAACTGGGTACTTATTTAGATGATCTTTATCG 600
DB	541	TATTGATCTATAGTAATATTTTACTAGAACTGGGTACTTATTTAGATGATCTTTATCG 600
QY	601	TATGATGGAAC-----AAATAC 618
DB	601	TATGATGGAACAAATACACATAATATCAATAAATAACATAATATCAATAAATAAC 660
QY	619	ACATAATATCAATAAGTTGCTTAAAGAAATTGCAAACTGGTTTAGCCAAACCAATCAGCAT 678
DB	661	ACATAATATCAATAAGTTGCTTAAAGAAATTGCAAACTGGTTTAGCCAAACCAATCAGCAT 720
QY	679	GTCCTATGTTAGTCGACCAATGTTGTAGGCAAAACGAGCGTTTCTGCTCGGTAGGAGG 738
DB	721	GTCCTATGTTAGTCGACCAATGTTGTAGGCAAAACGAGCGTTTCTGCTCGGTAGGAGG 780
QY	739	TTATAGAGATAAACTGCATTAGCCATTGGTGTGCGCTCAGCATTTACTGATCGCTTTAC 798
DB	781	TTATAGAGATAAACTGCATTAGCCATTGGTGTGCGCTCAGCATTTACTGATCGCTTTAC 840
QY	799	CGCTAAAGCGGGTGTAGCGTTCAATACCTAATGCGCGCATGCTTTATGCGTCTCTGT 858
DB	841	CGCTAAAGCGGGTGTAGCGTTCAATACCTAATGCGCGCATGCTTTATGCGTCTCTGT 900
QY	859	TGGTTATGAATTTCTAA 874
DB	901	TGGTTATGAATTTCTAA 916
RESULT 9		
AF187008		
LOCUS		

DEFINITION	Haemophilus ducreyi strain M90-02 serum resistance protein DsrA (dsrA) gene, complete cds.
ACCESSION	AF187008
VERSION	AF187008.1 GI:7188586
KEYWORDS	
SOURCE	Haemophilus ducreyi
ORGANISM	Haemophilus ducreyi Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales; Pasteurellaceae; Haemophilus.
REFERENCE	1 (bases 1 to 923) Elkins, C., Morrow, K.J. Jr. and Olsen, B. Serum resistance in Haemophilus ducreyi requires outer membrane protein DsrA Infect. Immun. 68 (3), 1608-1619 (2000)
AUTHORS	20143779
TITLE	10678980
JOURNAL	2 (bases 1 to 923) Elkins, C., Morrow, K.J. and Olsen, B. Direct Submission Submitted (17-SEP-1999) Division of Infectious Diseases, Department of Medicine, University of North Carolina at Chapel Hill, 521 Burnett-Womack Bldg. CB7030, Chapel Hill, NC 27516, USA
FEATURES	Location/Qualifiers 1..923 /organism="Haemophilus ducreyi" /mol_type="genomic DNA" /strain="M90-02" /db_xref="taxon:730" /note="serum resistant strain" 102..923 /gene="dsrA" 102..923 /gene="dsrA" /note="outer membrane protein required for serum resistance" /codon_start=1 /transl_table=11 /product="serum resistance protein DsrA" /protein_id="AAP37814.1" /db_xref="GI:7188587" /translation="MKIKLVAVVGLACSTITTTAAQPPKFAVSSLYSEYDYGKGL WTWSEGGDFDIKPGIKMKPKWISQATYLEQHYMPYTPVLVTSAPDVSPSSISIL LYPMSPDQDLGINROOLKLNLYSEYNDLRHDFKLKVLDAKISKNKONIDTISKVLEL GYLDGSRMMEQNTNHNKNTNHNKNTNHNKLSKELQTLGANSALSMVQPNQGV GKTSVAAGVGYRDKTALAIGVGRITDRFTAKGVAFNTNGMSYGAAGVGEF"
gene	
CDS	

ORIGIN		
Query Match	67.9%; Score 792.8; DB 1; Length 923;	
Best Local Similarity	94.0%; Pred. No. 3.4e-136;	
Matches	865; Conservative 0; Mismatches 7; Indels 48; Gaps 2;	
QY	3	AAATACGTCATTTGACATTTTATTAATGTAAGGTAGAATAAGAAAGTAAATCTATATTTA 62
DB	4	AATACGTCATTTGACATTTTATTAATGTAAGGTAGAATAAGAAAGTAAATCTATATTTA 63
QY	63	CAATCAAGATTGACAAATTAATTTACTTAATGAGGTGATTAAGAAATTAATGTTAGTTG 122
DB	64	CAATCAAGATTGACAAATTAATTTACTTAATGAGGTGATTAAGAAATTAATGTTAGTTG 123
QY	123	CCGTAGTGGGATTAGCTGTTCTTACTATTACAAATGGCTCAGCAGCCGCAAGTTTG 182
DB	124	CCGTAGTGGGATTAGCTGTTCTTACTATTACAAATGGCTCAGCAGCCGCAAGTTTG 183
QY	183	CTGGAGTATCTCTTTTGTATAGCTATGAGTATGATGTAAGGGTAAATGGAATGTTG 242
DB	184	CTGGAGTATCTCTTTTGTATAGCTATGAGTATGATGTAAGGGTAAATGGAATGTTG 243
QY	243	CTAATGAAGCGGTTTCGATATTAAGTGCAGGATTAATAATGAAGCCAAAGATGGA 302
DB	244	CTAATGAAGCGGTTTCGATATTAAGTGCAGGATTAATAATGAAGCCAAAGATGGA 303
QY	303	TTTCTAAACAGGCTACTTATCTTGAATTACAGATTATATGCTTACTCTGTTCTCG 362

Db	304	TTTCTAAACAGGCTACTTATCTTGAATTACAGCAATTATATGCTCTATATGCTCTCTGCTCG	363
Qy	363	TGACATATGCTCTCGGCTTCTCTAG-----CCCTATACCTGTATATCCGATGCTG	416
Db	364	TGACATCTGCTCTGAGCTTCTCTAGCTCTATCTCTATACTGTTATATCCGATGCTG	423
Qy	417	ATCTCGATCAACTTGGAAATAATCGGACGAGCTGAAATTTGAATTTGTATAGTTATTTTA	476
Db	424	ATCTCGATCAACTTGGAAATAATCGGACGAGCTGAAATTTGAATTTGTATAGTTATTTTA	483
Qy	477	ACGATTTAAGACACGATTTAAATTAAGATTTCTTGATGACAGTATTTCCAAAATAAAC	536
Db	484	ACGATTTAAGACACGATTTAAATTAAGATTTCTTGATGACAGTATTTCCAAAATAAAC	543
Qy	537	AAATATTTGATACATATAAGTAAATATTTACTAGAACTGGGTACTTATTTAGATGATCTT	596
Db	544	AAATATTTGATACATATAAGTAAATATTTACTAGAACTGGGTACTTATTTAGATGATCTT	603
Qy	597	ATCGTATGATGAAC-----AAA	614
Db	604	ATCGTATGATGAACAAAATACACATAATATCAATAAAATACACATAATATCAATAAAA	663
Qy	615	ATACACATAATATCAATAAGTTGTCTAAAGATTTGCAAACTGGTTAGCCCAACCAATCAG	674
Db	664	ATACACATAATATCAATAAGTTGTCTAAAGATTTGCAAACTGGTTAGCCCAACCAATCAG	723
Qy	675	CATTGCTATGTTAGTGCACCAATGCTGTAGGCAAAACGAGCGTTTCTGCTCGGTAG	734
Db	724	CATTGCTATGTTAGTGCACCAATGCTGTAGGCAAAACGAGCGTTTCTGCTCGGTAG	783
Qy	735	GAGGTTATAGAGATAAACTGCATTAGCCATTGGTGTGCGGTCAACGCAATTAAGTATGATCGCT	794
Db	784	GAGGTTATAGAGATAAACTGCATTAGCCATTGGTGTGCGGTCAACGCAATTAAGTATGATCGCT	843
Qy	795	TTACCGCTAAAGCGGTGTAGCGTTCAATACCTAACAATGCGGCAATGCTTTATGGTGCTT	854
Db	844	TTACCGCTAAAGCGGTGTAGCGTTCAATACCTAACAATGCGGCAATGCTTTATGGTGCTT	903
Qy	855	CTGTTGGTTATGAATTTCTAA 874	
Db	904	CTGTTGGTTATGAATTTCTAA 923	
RESULT 10			
AF187006			
LOCUS			
DEFINITION Haemophilus ducreyi strain CH1A serum resistance protein DsrA			
ACCESSION AF187006			
VERSION AF187006.1 GI:7188582			
KEYWORDS (dsrA) gene, complete cds.			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
FEATURES			
source			

/db_xref="taxon:730"			
/notes="intermediate serum susceptibility strain"			
103. .894			
/gene="dsrA"			
103. .894			
/gene="dsrA"			
/notes="outer membrane protein required for serum resistance"			
/codon_start=1			
/transl_table=11			
/product="serum resistance protein DsrA"			
/protein_id="AAF37812.1"			
/db_xref="GI:7188583"			
/translation="MKIKCLVAVVGLACSTITTTMAQPPKFAVGSSLDSEYDYDGKKG			
WWWSEKGDPIKAPGIMKPKWISQATVGLHYMPYTPVLVTYASAPNTVLLYP			
MPDOLGILNRQOLKILYSYFNDLRHGLKPLVLDARISONKONIDTISVYLKIGTY			
LSSYRMQNTNHNKNTNINKLSELOTLGANSALSMLVQPNVGKTSVSAAGV			
GYRDKTALAIGVGSRTIDRPTAKAGVAFNTYNGMSYGSVGYEP"			
ORIGIN			
Query Match			
Best Local Similarity 94.1%; Score 780.6; DB 1; Length 894;			
Matches 842; Conservative 0; Mismatches 29; Indels 24; Gaps 2;			
Qy	1	ATAAATACGTCATTCATTTTAAATGTAAGGTAGAATAAGAAAGTAAATTCATATT	60
Db	3	ATAAATACGTCATTCATTTTAAATGTAAGGTAAATACGAAAGTAAATTTATATT	62
Qy	61	TACAATCAAGATTGACAAATTTACTTAATGAGGTGATTATGAAATTTAAATGTTAGT	120
Db	63	TACAATCAAGATTGACAAATTTACTTAATGAGGTGATTATGAAATTTAAATGTTAGT	122
Qy	121	TGCGGTAGTGGGATTAGCTTGTCTACTATTACAACAATGGCTCAGCAGCGCCAAAGTT	180
Db	123	TGCGGTAGTGGGATTAGCTTGTCTACTATTACAACAATGGCTCAGCAGCGCCAAAGTT	182
Qy	181	TGCTGGAGTATCTTTTGTATAGTATGACTATGTAAGGGTAAATGGAGCTG	240
Db	183	TGCTGGAGTATCTTTTGTATAGTATGACTATGTAAGGGTAAATGGAGCTG	242
Qy	241	GTCTAATGAAGCGGTTTCGATATTAAAGTCCAGGGATTAAATGAAGCCAAAGATG	300
Db	243	GTCTGAAAAGACGGTTTCGATATTAAAGCGCCAGGGATTAAATGAAGCCAAAGATG	302
Qy	301	GATTTCTAAACAGCTACTTCTTGAATTTACAGCATTATATGCTTATCTCTGTTCT	360
Db	303	GATTTCTAGACAGCTACTTCTTGAATTTACAGCATTATATGCTTATCTCTGTTCT	362
Qy	361	CGTGACATATGCTCTGCGGCTTCTCTAGCCCTATATCTGTTATATCCGATGCTGATCC	420
Db	363	CGTGACATATGCTCTGCGGCTTCTCTAGCCCTATATCTGTTATATCCGATGCTGATCC	419
Qy	421	TGATCAACTTGAATAAATCGGACGAGCTGAAATTTGAATTTGATAGTTATTTTAACGA	480
Db	420	TGATCAACTTGAATAAATCGGACGAGCTGAAATTTGATAGTTATTTTAACGA	479
Qy	481	TTTAAAGACAGATTTTAAATTTAAAGTCTTGTATGACAGTATTTCCAAAATAACAAA	540
Db	480	TTTAAAGACAGATTTTAAATTTAAAGTCTTGTATGACAGTATTTCCAAAATAACAAA	539
Qy	541	TATTGATCTATTAAGTAAATTTTACTAGAACTGGGTACTTATTAGATGATTTCTTATCG	600
Db	540	TATTGATCTATTAAGTAAATTTTACTAGAACTGGGTACTTATTAGATGATTTCTTATCG	599
Qy	601	TATGATGGAAC-----AAAATACACATTAATCAATAGTTGTC	639
Db	600	TATGATGGAACAAAATAACACATAATATCAATAAATAATACACATAATATCAATAGTTGTC	659
Qy	640	TAAAGAAATTCAAACTGGTTTAGCCCAACCAATGCTGCTATGTTAGTGCAACCAAA	699
Db	660	TAAAGAAATTCAAACTGGTTTAGCCCAACCAATGCTGCTATGTTAGTGCAACCAAA	719
Qy	700	TGGTGTAGGCAAAACGAGCGCTTCTGCTGCGGTAGGAGGTTATAGAGATAAACTGCATT	759


```

Db      720 TGGTGTAGGCACAAACGAGCGTTTCTGCTGCGGTAGGAGGTTATAGAGTAAACGCGATT 779
Qy      760 AGCCATTGGTGTGCGGTCAAGCAATTACTGATCGCTTTACCGCTAAACGCGGTGTAGCGTT 819
Db      780 AGCCATTGGTGTGCGGTCAAGCAATTACTGATCGCTTTACCGCTAAACGCGGTGTAGCGTT 839
Qy      820 CAATACCTACAAATGGCGGCATGCTTATGCTGCTTCTGTTGTTATGAATTTCTAA 874
Db      840 CAATACCTACAAATGGCGGCATGCTTATGCTGCTTCTGTTGTTATGAATTTCTAA 894

RESULT 11
AB064943 1967 bp DNA linear BCT 24-MAY-2002
LOCUS
DEFINITION
Actinobacillus actinomycetemcomitans omp100 gene for outer membrane
protein 100, complete cds.
ACCESSION
AB064943
VERSION
AB064943.1 GI:19911153
KEYWORDS
SOURCE
ORGANISM
Actinobacillus actinomycetemcomitans
Actinobacillus actinomycetemcomitans
Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
Pasteurellaceae; Actinobacillus.
REFERENCE
1 Komatsuzawa,H., Asakawa,R., Kawai,T., Ochiai,K., Fujiwara,T.,
Taubman,M.A., Ohara,M., Kurihara,H. and Sugai,M.
Identification of six major outer membrane proteins from
Actinobacillus actinomycetemcomitans
JOURNAL
MEDLINE
Gene 288 (1-2), 195-201 (2002)
PUBMED
22030468
12034509
REFERENCE
2 (bases 1 to 1967)
Komatsuzawa,H.
AUTHORS
Direct Submission
TITLE
Submitted (04-JUL-2001) Hitoshi Komatsuzawa, Hiroshima University,
Faculty of Dentistry, Department of Microbiology; Minami-ku, Kasumi
1-2-3, Hiroshima 734-8553, Japan
JOURNAL
(E-mail:hkomatsu@hiroshima-u.ac.jp, Tel:81-82-257-5637,
Fax:81-82-25-75639)
FEATURES
source
Location/Qualifiers
1. .1967
/organism="Actinobacillus actinomycetemcomitans"
/mol_type="genomic DNA"
/db_xref="taxon:714"
602..1489
/gene="omp100"
602..1489
/gene="omp100"
/codon_start=1
/transl_table=11
/product="outer membrane protein 100"
/protein_id="BAB86905.1"
/db_xref="GI:19911154"
/translation="MTYQLFKHLVMTGALSVNALAKDSFLENPSANLPQVFN
RVDFINNETINENKDDIAINKANIASIEKDVNRNTGGIDRLAKQELVNRARITKNEL
DIRKNTKSTARIENKADINLEGNRVLQNDVDRSTENARSRANEQKIAENKKA
ENKADKADVENKADIAANSRAIRATSRSSQNTAALTTKVDRNTARIDRLSRNELD
KEVKNGLASQALSGLPQPNVNGSLNSAAVGVKSKTALAVGSYRFNQVNAKAGV
AVTNGGSAFYNVGLNPEW"
ORIGIN
Query Match 7.2%; Score 84.6; DB 1; Length 1967;
Best Local Similarity 51.5%; Pred. No. 7.9e-06;
Matches 195; Conservative 0; Mismatches 194; Indels 0; Gaps 0;
Qy 497 AAATTAAGTTCTTGATGACGATTTTCCAAAATAAACAATAATATGATCTATAGT 556
Db 1112 AAAACACAGACCGCATATTCGACGAAATTCACAGCGGATTGCAACCTTTAGATCTTCAAGC 1171
Qy 557 AAATATTACTAGACGGGTACTTTATTTAGATGATCTTATCGTATGATGGAACAAAT 616
Db 1172 CAAAACATCGCGCATTAACGACCAAAAGTTGATCGTAACTACTGCGCGTATGATCGATTA 1231

```

```

Qy 617 ACACATAATATCAATAGTTGTCTAAGAATTCCAACTCGTTTACCAACCAATCAGCA 676
Db 1232 GATAGCCGAGTCAATGAATTAGACAAGAGTAAAAAACCGTTTGGCTTCCCAAGCGGCA 1291
Qy 677 TTGCTATGTTAGTGCACCAACCAATGGTGTAGCGGAAAAACGAGCGTTTCTGCTGCGGTAGGA 736
Db 1292 CTAAGCGGCTTATCCAAACGATATATGTCGGCAGCCTTAACCTTGAGTGACGCTGTTGGT 1351
Qy 737 GGTTATAGAGATAAACTGCAATAGCCATTGGTGTGCGCTCACGCAATCTATGATCGCTTT 796
Db 1352 GGTATAAATCTAAAAACAGCACATAGCGGTTGGTTTCAGGTTATCGTTTCAATCAAAATGTA 1411
Qy 797 ACCGCTAAAGCGGTGTAGCGTTCAATACCTACCAATGGCGGCATGCTTATGCTGCTTCT 856
Db 1412 GCCCGAAGCGCGGTGTGGCAGTAAGTACCAATGGTGGCAGCGCAACTATAACGTCGGT 1471
Qy 857 GTTGGTTATGAATTTCTAAAT 875
Db 1472 TTAACTTTGAGTGGTAAAT 1490

RESULT 12
AF316502 2579 bp DNA linear BCT 21-MAR-2002
LOCUS
DEFINITION
Actinobacillus actinomycetemcomitans putative topoisomerase IV
subunit B (Aa32-1-1) gene, partial cds; and putative
adhesin/invasin (Aa32-1-2) gene, complete cds.
ACCESSION
AF316502
VERSION
AF316502.1 GI:19568162
KEYWORDS
SOURCE
ORGANISM
Actinobacillus actinomycetemcomitans
Actinobacillus actinomycetemcomitans
Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
Pasteurellaceae; Actinobacillus.
REFERENCE
1 (bases 1 to 2579)
Lepine,G., Li,L. and Ellen,R.P.
Cloning and characterization of three invasive genes of
Actinobacillus actinomycetemcomitans
Unpublished
2 (bases 1 to 2579)
Lepine,G., Li,L. and Ellen,R.P.
Direct Submission
AUTHORS
Submitted (25-OCT-2000) Dental Research Institute, University of
Toronto, 124 Edward St., Room 442, Toronto, Ontario M5G 1G6, Canada
JOURNAL
Location/Qualifiers
FEATURES
source
1. .2579
/organism="Actinobacillus actinomycetemcomitans"
/mol_type="genomic DNA"
/strain="UT32"
/db_xref="taxon:714"
complement(<1..882)
/gene="Aa32-1-1"
complement(<1..882)
/gene="Aa32-1-1"
/notes="similar to Haemophilus influenzae topoisomerase IV
subunit B protein deposited in GenBank Accession Number
AAC23174"
/transl_table=11
/codon_start=1
/product="putative topoisomerase IV subunit B"
/protein_id="AAL91672.1"
/db_xref="GI:19568163"
/translation="MTNTYSAEITVLKDLPEVQLRPGMYTDTTRPNHLAQEVINDNSV
DRLACGAFATKIEVILHKDQSIIEVDNRMGPVDIHPVEKVGSGVEVIMSKLHAGKFSN
KNYTFSGGLHGVGISVUNALSERVDYTKRNSIYKIAPENGKKVDLTWTCGRPT
TGTVTHFKPKGYTFDSEKFSRKLRLHLLKRAKVLCSGLEIKFIDNVDNDEDTWLYQDG
LNDLYMEAVNGLVTLPEQPFIFGEFNGEKAWSWALLPPEGGSLIGESVYNNLIPTALG
GTHVNLGRQLLDAMREF"
1385..2272
/gene="Aa32-1-2"
1385..2272
/gene="Aa32-1-2"

```

/note="similar to Neisseria meningitidis MC58
adhesin/invasin deposited in GenBank Accession Number
AAF42321 and Moraxella catarrhalis ubiquitous surface
protein A2 deposited in GenBank Accession Number AAB96391"
/codon_start=1
/transl_table=11
/product="putative adhesin/invasin"
/protein_id="AAL91673.1"
/db_xref="GI:19568164"
/translation="MTYLPKHLVALMTVGAISVNALAKDSFLENPSANLPQVFNQ
RVDIPNNETHINENKDIANKANIASIEKDVMTGCDRLAKQELVNRATYKREL
DIQNTKISIAENTADIGNLEGVNRVLQNDVVRSTENAAKSRANEQKIAENKKAI
ENKADKADVENKADIAANGRAIATFFSSQNTAALTTKVDRTARIDRLDSRVNELD
KEVKNGLASQAALSGLPQPNVNSLNSAAVGGYKSKTALAVALGSGYRFNQVAAKAGV
AVSTNGGSATVNVGLNFEW"

ORIGIN

Query Match 7.2%; Score 84.6; DB 1; Length 2579;
Best Local Similarity 51.5%; Pred. No. 7.4e-06;
Matches 195; Conservative 0; Mismatches 184; Indels 0; Gaps 0;

Qy 497 AAATTAAAGTCTTGATGCGAGTATTCCAAAATAAACAATAATTGACTATAAGT 556
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1895 AAAACAGAGCGGATATTGCAGAAATCCAGAGGATTGCAACTTTAGATCTCAAGC 1954
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 557 AAATATTACTAGAACTGGGTACTTATTATTGATGATCTTATCGTATGATGGAACAAAAT 616
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1955 CAAAACATCGCGGCATTAAACGACCAAGTTGATCGTAATCTGCGGTATTGATCGATTA 2014
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 617 ACACATATATCAATAGTTGTTCTTAAGAATTGCAAACTGGTTTAGCGCAACCAATACGA 676
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 2015 GATAGCGGAGTCAATGAATTAGACAAAGAGTAAACAAACGGTTTGGCTTCCCAAGCGCA 2074
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 677 TTGCTATGTTAGTGCACCAAAATGGTGTAGGCAAAACGCGTTTCTGCTCGCGTAGGA 736
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 2075 CTAAGCGGCTTATTCCAAACCGTATAATGTGCGCAGCCTTAACCTTGAGTGCAGCTGTTGGT 2134
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 737 GGTATTAGAGATAAACTGCAATTAGCCATTGGTGTGCGGCTCAGCCATTACTGATCGCTTT 796
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 2135 GGTATAAATCTAAACAGCACTAGCGTTGGTTTCAGGTATCGTTTCAATCAAAATGTA 2194
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 797 ACCGCTAAGCGGTGTAGCGTTCAATACCTACAAATGCGCGCATGTCTTATGGTCTTCT 856
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 2195 GCCGGAAGGCGGTGTGCGAGTAGTACCAATGTTGCGCAGCAACCTATAACGTCGGT 2254
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 857 GTTGGTTATGAATCTTAAT 875
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 2255 TTAAACTTTGATGGTGAAT 2273
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 13
YPYOPA YPYOPA 2216 bp DNA linear BCT 12-SEP-1993
LOCUS Yersinia pseudotuberculosis virulence plasmid pIBI yopA gene for
DEFINITION YopI protein.
ACCESSION X13883 X12758
VERSION X13883.1 GI:48639
KEYWORDS invasin; plasmid; yopA gene.
SOURCE Yersinia pseudotuberculosis
ORGANISM Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Yersinia.
1 (bases 357 to 1729)
Rosqvist,R.; Skurnik,M. and Wolf-Watz,H.
REFERENCES Increased virulence of Yersinia pseudotuberculosis by two
AUTHORS independent mutations
TITLE Nature 334 (6182), 522-524 (1988)
JOURNAL 88302441
MEDLINE 3043229
PUBMED
REFERENCE 2 (bases 1 to 2216)
AUTHORS Skurnik,M. and Wolf-Watz,H.
TITLE Analysis of the yopA gene encoding the YopI virulence determinants
of Yersinia spp

JOURNAL Mol. Microbiol. 3 (4), 517-529 (1989)
MEDLINE 89343638
PUBMED 2761389
REFERENCE 3 (bases 1 to 2216)
AUTHORS Skurnik,M.
TITLE Direct Submission
JOURNAL Submitted (30-DEC-1988) Skurnik M., Department of Medical
COMMENT Microbiology, University of Turku, SF-20520 Turku, Finland
FEATURES Data kindly reviewed (05-JUL-1989) by Skurnik M.
source Location/Qualifiers
1..2216
/organism="Yersinia pseudotuberculosis"
/mol_type="genomic DNA"
/strain="ypIII"
/db_xref="taxon:633"
/clone="pYMS 2, pIBI540"
/plasmid="pIBI"
54..82
promoter /notes="promoter region"
88..1720
precursor_RNA /notes="yopA transcript"
357..1661
CDS /note="unnamed protein product; YopI preprotein (AA 1 - 434)"
/codon_start=1
/transl_table=11
/protein_id="CAA32088.1"
/db_xref="GI:48640"
/db_xref="GOA:P10858"
/db_xref="SWISS-PROT:P10858"
/translation="MTKDFKISVSAALISALFSPYPFAPEPEDNGDIPRLSAVQIS
PNVDPKLVGLYPAKPTILROENKPLPRGPQPEKKRLAEAIQPOVLGAGGLNARA
KDPIYTAIGATAEAAKPAAVAVGSGTATGNSVAIGPLSKALGDSAVTVYGASSTAK
DGVATGARASADTGVAVGNSKVDACNSVAIGHSSHVADHGYSIAIGHSKTDREN
SVSIGHESLNRLTHLAAGTETDANVAQLKEMARETLENARKETLAQSDNVLDAAK
KHNSVQYTTDHFESQLDNRDLKDRKGLASSAALNSLFPQYGVGVKNVATGVGGY
AISESNQITDHFESQLDNRDLKDRKGLASSAALNSLFPQYGVGVKNVATGVGGY
RSSQALAIQSGYRVNESVALKAGVAVAGSSNMVYNASFNIEW"
1695..1720
terminator /note="transcription terminator"
ORIGIN

Query Match 5.9%; Score 68.6; DB 1; Length 2216;
Best Local Similarity 54.6%; Pred. No. 0.0067;
Matches 137; Conservative 0; Mismatches 114; Indels 0; Gaps 0;

Qy 568 AGAAGTGGTACTATTATGATGATCTTATCGTATGATGGAACAAATACACATAAT 627
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1358 AGCAATCAGTAATCTAATCAATACACAGATCAATAATCAGTCAACTTGACACCGTTT 1417
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 628 CAATAAGTTCTTAAAGAAATTGCAAACTGGTTTAGCCAAACCAATCAGCATTTCTATGTT 687
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1418 AGATAAATTGACAAACGAGTTGCAAAAGTTTAGCCAGTTTCAGCCGCTTTAAACAGCTT 1477
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 688 AGTGCAACCAAAATGGTGTAGGCAAAACGAGCGTTTCTGCTGCGGTAGAGTTTATAGAGA 747
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1478 GTTCCAGCCATATGGTGTAGGAAAGTAAACTTTACTGCGAGTGTGCGGGGATATCGTTC 1537
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 748 TAAACTGCAATTAGCCATTCGTGTCGCTCAGCATTTACTGATCGCTTACCGGTAAGC 807
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1538 TAGTCAGGCATTAGCAATTTGCTTCTGCTATCGTGTAAATGAGAGTGTGCGCACTTAAAGC 1597
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 808 GGGTGTAGCGT 818
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1598 CGGTGTGCGCTT 1608
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 14
YPYOPANF YPVOPANF 2216 bp DNA linear BCT 01-DEC-1992
LOCUS Yersinia pestis virulence plasmid pYV019 yopA pseudogene.
DEFINITION Yersinia pestis virulence plasmid pYV019 yopA pseudogene.
ACCESSION X13880 X12759
VERSION X13880.1 GI:48641

KEYWORDS plasmid; pseudogene; putative adhesin; yopA gene.

SOURCE Yersinia pestis

ORGANISM Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Yersinia.

REFERENCE 1 (bases 307 to 1678)

AUTHORS Rosqvist, R., Skurnik, M. and Wolf-Watz, H.

TITLE Increased virulence of Yersinia pseudotuberculosis by two independent mutations

JOURNAL Nature 334 (6182), 522-524 (1988)

MEDLINE 88302441

PUBMED 3043229

REFERENCE 2 (bases 1 to 2216)

AUTHORS Skurnik, M. and Wolf-Watz, H.

TITLE Analysis of the yopA gene encoding the YopI virulence determinants of Yersinia spp

JOURNAL Mol. Microbiol. 3 (4), 517-529 (1989)

MEDLINE 89343638

PUBMED 2761389

REFERENCE 3 (bases 1 to 2216)

AUTHORS Skurnik, M.

TITLE Direct Submission

JOURNAL Submitted (30-DEC-1988) Skurnik M., Department of Medical Microbiology, University of Turku, SF-20520 Turku, Finland

COMMENT see X13883 for functional Y. pseudotuberculosis yopA gene. Data kindly reviewed (05-jul-1989) by Skurnik M.

FEATURES

source

1. .2216

/organism="Yersinia pestis"

/mol_type="genomic DNA"

/strain="019"

/db_xref="taxon:632"

/clone="pYMS 1, pYMS 1080."

/plasmid="pYV019"

4. .32

/note="promoter region"

307

/note="yopA translation start (Y. pseudotuberculosis)"

539

/note="deletion site"

1643. .1669

/note="transcription terminator"

ORIGIN

Query Match 5.9%; Score 68.4; DB 1; Length 2216;

Best Local Similarity 56.8%; Pred. No. 0.0073;

Matches 126; Conservative 0; Mismatches 96; Indels 0; Gaps 0;

Qy 597 ATCGTATGATGGACAAATACACATAATATCATATAGTTGCTTAAGAAATTCGAAACTG 656

Db 1336 ATCATAAATTCAGTCAACTTGACAAACCGTTTAGATAACTTGACAAACGATGACAAAG 1395

Qy 657 GTTTAGCCAAACCAATCAGCATTTGCTATGTTAGTGCAACAAATGGTGTAGGCAAAACGA 716

Db 1396 GTTTAGCCAGTTGAGCGCTTTAAACAGCTTTGTCAGCCATATGTTGTAGGGAAGTAA 1455

Qy 717 GCGTTTCTGTCGGTAGGAGGTATATAGATAAACTGCATTAACCATGCGTGTGCGGT 776

Db 1456 ACTTTCCTCGAGGTTGCGGGGATATCGTTCTAGTCAGGCATTAGCAATGTTCTGCTT 1515

Qy 777 CAGCATTAATGATCGCTTTTACCGCTAAAGCGGTGTAGCGT 818

Db 1516 ATCGTGTAAATGAGAGTGTGCGCATTTAAAGCGGTGTGCGT 1557

RESULT 15

LOCUS YPCD1 70305 bp DNA circular BCT 18-JUN-2003

DEFINITION Yersinia pestis plasmid pCD1.

ACCESSION AL117189

VERSION AL117189.1 GI:5832423

KEYWORDS chaperone; cytotoxic effector; IS100; IS1616; IS1617; lcr; low-calcium response; syc; targeted effector; transposase; type III

secretion; V antigen; virulence; ylp; yop; ysc.

Yersinia pestis

Yersinia pestis

Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Yersinia.

REFERENCE 1 (bases 1 to 70305)

AUTHORS Parkhill, J., Wren, B.W., Thomson, N.R., Titball, R.W., Holden, M.T.G., Prentice, M.B., Sebahia, M., James, K.D., Churcher, C., Mungall, K.L., Baker, S., Basham, D., Bentley, S.D., Brooks, K., Cerdeno-Tarraga, A.M., Chillingworth, T., Cronin, A., Davies, R.M., Davis, P., Dougan, G., Feltwell, T., Hamlin, N., Holroyd, S., Jagels, K., Leather, S., Karlyshev, A.V., Moule, S., Oyston, P.C.F., Quail, M., Rutherford, K., Simmonds, M., Skelton, J., Stevens, K., Whitehead, S. and Barrall, B.G.

TITLE Genome sequence of Yersinia pestis, the causative agent of plague

JOURNAL Nature 413 (6855), 523-527 (2001)

MEDLINE 21470413

PUBMED 11586360

REFERENCE 2 (bases 1 to 70305)

AUTHORS James, K.D., Parkhill, J., Barrall, B.G. and Rajandream, M.A.

TITLE Direct Submission

JOURNAL Submitted (03-SEP-1999) Yersinia pestis sequencing project, The Wellcome Trust Sanger Institute, Genome Campus, Hinxton, Cambridge CB10 1SA E-mail: parkhill@sanger.ac.uk DNA supplied by Dr. Andrew Karlyshev and Prof. Brendan Wren, [3]. Department of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT

Notes:

Details of Y. pestis sequencing at the Sanger Centre are available on the World Wide Web.

(URL, http://www.sanger.ac.uk/Projects/Y_pestis/).

FEATURES

source

1. .70305

/organism="Yersinia pestis"

/mol_type="genomic DNA"

/strain="CO-92 Biovar Orientalis"

/db_xref="taxon:632"

/plasmid="pCD1"

1. .1356

/note="IS100 element"

88. .1110

/genes="YPCD1.01"

88. .1110

/gene="YPCD1.01"

/note="YPCD1.01, probable transposase, len: 340 aa; putative insertion sequence IS100, identical to corresponding CDS from Y.pestis KIM5 pCD1 (EMBL:AF053946, AF074612) (340 aa), fasta scores; Opt: 2328 z-score: 3808.9 E(): 0, 100.0% identity in 340 aa overlap. Similar to many others e.g. TRAO_ECOLI (EMBL:X14793), istA, E.coli transposase for insertion sequence element IS21 (390 aa) (33.1% identity in 329 aa overlap). Contains Pflam match to entry PF00239 recombinase, Site-specific recombinases, score 25.70, E-value 4.8e-06. Contains probable helix-turn-helix motif at aa 19-40 (Score 2045, +6.15 SD)"

/codon_start=1

/transl_table=11

/product="putative transposase"

/protein_id="CAB54878.1"

/db_xref="GI:5832424"

/db_xref="GOA:P74993"

/db_xref="SPTREMBL:P74993"

/translation="MVTFETVMEIKLHKQMSRAIARELGRNTRVRYLQKSP PKYTPRPAVASLLDEYRIQRIADAHPIKATVAREIQDQYRGGMVILRAFIR SLSPQEQPAPAFETEPGRQMDVWNGRSPFLVFVAVLGSRMLYIEFTDNMR YDTLETCHRNAPFRFGVPREVLYDNMKTVLORDAYQTQHRFHPSLWQKMGMPFS PRLCPFPRAQTKGKVERMYQYTNSEFVPLMTFLRPMGIVDVETANRHLRLHLDVA NQRKHETIQARPCDRWLEEQSQSMALPPEKKEYDVHLDENVFDRKPLHPLSIYDS FCRGVA"

112. .195

/genes="YPCD1.01"

/note="Pflam match to entry PF00239 recombinase, Site-specific recombinases, score 25.70, E-value 4.8e-06"

1110. .1889

misc_feature

gene

DE Amino acid sequence of the DsrA protein from strain CIPA77.
 XX DsrA protein; outer membrane protein; serum resistance; vaccine;
 KW chancroid disease; genital ulcer disease.
 XX Haemophilus ducreyi.
 XX WO200104138-A1.
 XX 18-JAN-2001.
 XX 07-JUL-2000; 2000WO-US018834.
 XX 09-JUL-1999; 99US-0143257P.
 XX (YPNC-) UNIV NORTH CAROLINA.
 PA (ELKI//) ELKINS C.
 XX Elkins C;
 XX WPI; 2001-138311/14.
 DR N-PSDB; AAF25264.
 XX Novel purified DsrA protein of Haemophilus ducreyi which confers serum
 PT resistance to the bacteria used to produce vaccines that induce immune
 PT response against the bacteria in subject at risk of developing chancroid.
 XX Claim 7; Page 54; 80pp; English.
 XX The present sequence represents a DsrA 30 kDa protein of Haemophilus
 CC ducreyi. DsrA is an outer membrane protein that confers serum resistance
 CC to the bacteria. DsrA antisense oligonucleotides are useful for detecting
 CC a polynucleotide which encodes DsrA in a biological sample. The DsrA
 CC polypeptide is used to produce vaccine compositions, which are useful for
 CC inducing a protective immune response in a subject at risk of developing
 CC Haemophilus ducreyi infection e.g. chancroid, a genital ulcer disease
 CC transmitted by sexual contact. DsrA, its catalytic or immunogenic
 CC fragments or peptides can be used for screening libraries of compounds in
 CC a variety of drug screening techniques. The proteins and peptides may be
 CC used as antigens in immunoassays for detection of Haemophilus ducreyi in
 CC various tissues and body fluids, e.g. blood, spinal fluid, sputum etc. .
 CC The nucleic acids are useful for the preparation of DsrA proteins
 XX SQ Sequence 264 AA;
 Alignment Scores:
 Pred. No.: 6.83e-143 Length: 264
 Score: 1299.50 Matches: 253
 Percent Similarity: 95.83% Conservative: 0
 Best Local Similarity: 95.83% Mismatches: 4
 Query Match: 63.45% Indels: 7
 DB: 4 Gaps: 1
 US-10-030-529A-1 (1-1168) x AAB31707 (1-264)
 QY 101 ATGAAATTAAGTCTAGTTCGCGTAGTGGAGTACCTGTTCTACTATTACAAATG 160
 Db 1 MethylyleCysLeuValAlaValAlaValGlyLeuAlaCysSerThrIleThrMet 20
 QY 161 GCTCAGCAGCGCCAAAGATTTGCTGGAGTATCTTTTGTATAGTATGATGACTAT 220
 Db 21 AlaGlnGlnProProlysPheAlaGlyValSerSerLeuTySerTyGluTyAspIyr 40
 QY 221 GGTAAGGGTAAATGGACTTGGTCTTAATCAAGCGCGGTTTCGATATTAAGTCCAGGGATT 280
 Db 41 GlylysGlyLysTrpThrTrpSerAsnGluGlyGlyPheAspIleLysValProGlyIle 60
 QY 281 AAATGAAGCCAAAGAAATGGATTCTTAACAGGCTACTTATCTTGAATTACAGCATAT 340
 Db 61 LysMetLysProlysGluTrpIleSerLysGlnAlaThrTyLeuGluLeuGlnHisTy 80
 QY 341 ATGCCTATATCTCTGTTCTCGTACATATGCTCTCGCGCTTCTCTAGCCCTATCTG 400
 Db 81 MetProTyThrProValLeuValThrTyAlaHisAspValProProSerSerIleLeu 100
 QY 401 TTATATCCGATGTCTGATCTGATCAACTTGGATTAATAATCGGACGAGCTGAATTTGAAT 460
 Db 101 LeuTyProMetSerAspProAspGlnLeuGlyIleAsnArgGlnGlnLeuLysLeuAsn 120
 QY 461 TTGTATAGTTATTTAAACGATTTAAAGACACGATTTTAAATTTAAAGTTCTTTGATGCACGT 520
 Db 121 LeuTySerTyThrPheAsnAspLeuArgHisAspPheLysLeuLysValLeuAspAlaArg 140
 QY 521 ATTTCCAAAAATAACAAAAATATTGATCTATATAAGTAAATATTATTACTAGAACTGGGTACT 580
 Db 141 IleSerLysAsnLysGlnAsnIleAspThrIleSerLysTyLeuLeuGluLeuGlyThr 160
 QY 581 TATTAGATGATTTCTTATCGTATGATGAACA-----AATACA 619
 Db 161 TyLeuAspAspSerTyArgMetMetGluGlnAsnThrHisAsnIleAsnLysAsnThr 180
 QY 620 CATAATATCAATAAGTTGTCTAAAGAAATTCGAACTGGTTTAGCAACCAATCAGCATTTG 679
 Db 181 HisAsnIleAsnLysLeuSerLysGluLeuGlnThrGlyLeuAlaAsnGlnSerAlaLeu 200
 QY 680 TCTATGTTAGTCAACCAAAATGGTGTAGGCAAAACGAGCGTTTCTGCTGCGGTAGGAGGT 739
 Db 201 SerMetLeuValGlnProAsnGlyValGlyLysThrSerValSerAlaAlaValGlyGly 220
 QY 740 TATAGAGATAAACTGCATTTAGCATTTGGTGTGGCTCAGCATTTACTGATCGCTTTACC 799
 Db 221 TyArgAspLysThrAlaLeuAlaIleGlyValGlySerArgIleThrAspArgPheThr 240
 QY 800 GCTAAAGCGGTGTAGCGTTCAATACCTACAAATGGCGGCATGCTTATGGTCTTCTGTT 859
 Db 241 AlaLysAlaGlyValAlaPheAsnThrTyAsnGlyGlyMetSerTyGlyAlaSerVal 260
 QY 860 GGTATGAAATTC 871
 Db 261 GlyTyGluPhe 264
 RESULT 6
 AAB31711
 ID AAB31711 standard; protein; 271 AA.
 XX AAB31711;
 XX 30-APR-2001 (first entry)
 DE Amino acid sequence of the DsrA protein from strain V-1157.
 XX DsrA protein; outer membrane protein; serum resistance; vaccine;
 KW chancroid disease; genital ulcer disease.
 XX Haemophilus ducreyi.
 XX WO200104138-A1.
 XX 18-JAN-2001.
 XX 07-JUL-2000; 2000WO-US018834.
 XX 09-JUL-1999; 99US-0143257P.
 XX (YPNC-) UNIV NORTH CAROLINA.
 PA (ELKI//) ELKINS C.
 XX Elkins C;
 XX WPI; 2001-138311/14.
 DR N-PSDB; AAF25268.
 XX Novel purified DsrA protein of Haemophilus ducreyi which confers serum
 PT resistance to the bacteria used to produce vaccines that induce immune
 PT response against the bacteria in subject at risk of developing chancroid.
 XX

```
PS Claim 7; Page 57; 80pp; English.
XX
CC The present sequence represents a DsrA 30 kDa protein of Haemophilus
CC ducreyi. DsrA is an outer membrane protein that confers serum resistance
CC to the bacteria. DsrA antisense oligonucleotides are useful for detecting
CC a polynucleotide which encodes DsrA in a biological sample. The DsrA
CC polypeptide is used to produce vaccine compositions, which are useful for
CC inducing a protective immune response in a subject at risk of developing
CC Haemophilus ducreyi infection e.g. chancroid, a genital ulcer disease
CC transmitted by sexual contact. DsrA, its catalytic or immunogenic
CC fragments or peptides can be used for screening libraries of compounds in
CC a variety of drug screening techniques. The proteins and peptides may be
CC used as antigens in immunoassays for detection of Haemophilus ducreyi in
CC various tissues and body fluids, e.g. blood, spinal fluid, sputum etc.
CC The nucleic acids are useful for the preparation of DsrA proteins
XX
SQ Sequence 271 AA;

Alignment Scores:
Pred. No.: 1.18e-141 Length: 271
Score: 1289.00 Matches: 252
Percent Similarity: 92.99% Conservative: 0
Best Local Similarity: 92.99% Mismatches: 5
Query Match: 62.94% Indels: 14
DB: 4 Gaps: 1

US-10-030-529A-1 (1-1168) x AAB31711 (1-271)

Qy 101 ATGAAATTAATGTTAGTTCGCGTAGTGGATTAGCTTCTTCTACTATTACAAATG 160
Db 1 MetLysIleLysCysLeuValAlaValGlyLeuAlaCysSerThrIleThrMet 20

Qy 161 GCTCAGCAGCGCCAAAGATTGCTGGAGTACTCTTTGTATAGCTATGATGACTAT 220
Db 21 AlaGlnGlnProProlLysPheAlaGlyValSerSerLeuTyrSerTyrGluTyrAspTyr 40

Qy 221 GGTAAGGTTAAATGGACTTGTCTTAATGAAGCGGTTTCGATATATAAGTCCAGGATT 280
Db 41 GlyLysGlyLysTyrThrTrpSerAsnGluGlyGlyPheAspIleLysValProGlyIle 60

Qy 281 AAATGAAGCCAAAGATGATTCTTAACAGCGTACTTATCTTGATTTACAGCATAT 340
Db 61 LysMetLysProLysGluTrpIleSerLysGlnAlaThrTyrLeuGluLeuGlnHisTyr 80

Qy 341 ATGCTTATATCTCTGTCGATACATATCTCTCGCGTTTCTCTAGCCCTTACTG 400
Db 81 MetBroTyrThrProValLeuValThrSerAlaProAspValProSerSerIleLeu 100

Qy 401 TTATATCCGATGTCGTCTGATCACTTGAATAAATCGCAGCAGCTGAAATTTGAAT 460
Db 101 LeuTyrProMetSerAspProAspGlnLeuGlyIleAsnArgGlnGlnLeuLysLeuAsn 120

Qy 461 TTGTATAGTTATTTAAGCATTTAAGACACAGTTTAAATTAAGTTCTTGATGACGT 520
Db 121 LeuTyrSerTyrPheAsnAspLeuArgHisAspPheLysLeuLysValLeuAspAlaArg 140

Qy 521 ATTTCCAAATAACAAATATTGATATAAGTAAATATTACTAGACTGGTACT 580
Db 141 IleSerLysAsnLysGlnAsnIleAspThrIleSerLysTyrLeuLeuGluLeuGlyThr 160

Qy 581 TATTAGATGATTCTTATCGTATGATGAAACAA----- 613
Db 161 TyrLeuAspGlySerTyrArgMetMetGluGlnAsnThrHisAsnIleAsnLysAsnThr 180

Qy 614 -----AATACATATAATCATTAATAGTTGCTCTAAAGATTTCGAACCTGGT 658
Db 181 HisAsnIleAsnLysAsnThrHisAsnIleAsnLysLeuSerLysGluLeuGlnThrGly 200

Qy 659 TTAGCCCAACCAATCAGCATTCCTATGTTAGTGCACCAACCAATGTTAGGCAACACGAGC 718
Db 201 LeuAlaAsnGlnSerAlaLeuSerMetLeuValGlnProAsnGlyValGlyLysThrSer 220

Qy 719 GTTTCTGCTGCGTAGGAGGTTATAGAGATAAAACTGCATTAGCCATTGGTGTGCGGTCA 778
```

```

Db 221 ValSerAlaAlaValGlyGlyTyrArgAspLysThrAlaLeuAlaIleGlyValGlySer 240
Qy 779 CGCATTACTGATCGCTTACCGCTAAAGCGGGTGTAGCGTTCAATACCTCAATGGCGGC 838
Db 241 ArgIleThrAspArgPheThrAlaLysAlaGlyValAlaPheAsnThrTyrAsnGlyGly 260
Qy 839 ATGCTTATGCTGCTCTCTGTTGTTGTTATGAATTC 871
Db 261 MetSerTyrGlyAlaSerValGlyTyrGluPhe 271

RESULT 7
AAB31712
ID AAB31712 standard; protein; 273 AA.
XX
AC AAB31712;
XX
DT 30-APR-2001 (first entry)
XX
DE Amino acid sequence of the DsrA protein from strain M90-02.
XX
KW DsrA protein; outer membrane protein; serum resistance; vaccine;
KW chancroid disease; genital ulcer disease.
XX
OS Haemophilus ducreyi.
XX
PN WO200104138-A1.
XX
PD 18-JAN-2001.
XX
PF 07-JUL-2000; 2000WO-US018834.
XX
PR 09-JUL-1999; 99US-0143257P.
XX
PA (UYN-) UNIV NORTH CAROLINA.
PA (ELKI) ELKINS C.
XX
PI Elkins C;
XX
WP: 2001-138311/14.
XX
N-PSDB; AAF25269.
XX
Novel purified DsrA protein of Haemophilus ducreyi which confers serum
resistance to the bacteria used to produce vaccines that induce immune
response against the bacteria in subject at risk of developing chancroid.
Claim 7; Page 58; 80pp; English.
XX
CC The present sequence represents a DsrA 30 kDa protein of Haemophilus
CC ducreyi. DsrA is an outer membrane protein that confers serum resistance
CC to the bacteria. DsrA antisense oligonucleotides are useful for detecting
CC a polynucleotide which encodes DsrA in a biological sample. The DsrA
CC polypeptide is used to produce vaccine compositions, which are useful for
CC inducing a protective immune response in a subject at risk of developing
CC Haemophilus ducreyi infection e.g. chancroid, a genital ulcer disease
CC transmitted by sexual contact. DsrA, its catalytic or immunogenic
CC fragments or peptides can be used for screening libraries of compounds in
CC a variety of drug screening techniques. The proteins and peptides may be
CC used as antigens in immunoassays for detection of Haemophilus ducreyi in
CC various tissues and body fluids, e.g. blood, spinal fluid, sputum etc.
CC The nucleic acids are useful for the preparation of DsrA proteins
XX
SQ Sequence 273 AA;

Alignment Scores:
Pred. No.: 5.96e-141 Length: 273
Score: 1283.00 Matches: 253
Percent Similarity: 92.67% Conservative: 0
Best Local Similarity: 92.67% Mismatches: 4
Query Match: 62.65% Indels: 16
DB: 4 Gaps: 2

US-10-030-529A-1 (1-1168) x AAB31712 (1-273)
```

```
QY 101 ATGAATAAATGCTTTAGTTGCGTAGTGGATTAGCTTGTCTTACTATTACAAATG 160
Db |||||||
QY 1 MetLysIleLysCysLeuValAlaValAlaValGlyLeuAlaCysSerThrIleThrMet 20
Db |||||||
QY 161 GCTCAGCAGCGCCCAAAAGTTGCTGGAGTATCTTTTGTATAGCTATGAGTATGACTAT 220
Db |||||||
QY 21 AlaGlnGlnProProLysPheAlaGlyValSerSerLeuTyrSerTyrGluTyrAspTyr 40
Db |||||||
QY 221 GGTAAAGGTTAAATGGACTGTCTTAATGAAGCGCGTTTCGATATTAAGTCCAGGATT 280
Db |||||||
QY 41 GlyLysGlyLysTrpThrTrpSerAsnGluGlyGlyPheAspIleLysValProGlyIle 60
Db |||||||
QY 281 AAAATGAAGCCAAAGAAATGATTTCTAAACAGCTACTTATCTTGAATTACAGCATAT 340
Db |||||||
QY 61 LysMetLysProLysGluTrpIleSerLysGlnAlaThrTyrLeuGluLeuGlnHisTyr 80
Db |||||||
QY 341 ATGCTTATATCTCTGTTCTGTCGACATATGCTCTCGCGGTTTCTCTAGC-----CCT 394
Db |||||||
QY 81 MetProTyrThrProValLeuValThrSerAlaProAspValSerProSerSerIleSer 100
Db |||||||
QY 395 ATACTGTATATCCGATGTCGATCCGATCAACTGGAATAAATCGCAGAGCTGAAA 454
Db |||||||
QY 101 IleLeuLeuTyrProMetSerAspProAspGlnLeuGlyIleAsnArgGlnLeuLys 120
Db |||||||
QY 455 TTGAATTGTATAGTTATTTTAAAGATTAAAGACACGATTTTAAATTTAAAGTTCTTGAT 514
Db |||||||
QY 121 LeuAsnLeuTyrSerTyrPheAsnAspLeuArgHisAspPheLysLeuLysValLeuAsp 140
Db |||||||
QY 515 GCAGCTATTTCCAAAAATAACAAATATTGATATAAGTAAATATTACTAGAACTG 574
Db |||||||
QY 141 AlaArgIleSerLysAsnLysGlnAsnIleAspThrIleSerLysTyrLeuLeuGluLeu 160
Db |||||||
QY 575 GGTACTTATTAGATGATTTCTTATCGTATGATGAACAA----- 613
Db |||||||
QY 161 GlyThrTyrLeuAspGlySerTyrArgMetMetGluGlnAsnThrHisAsnIleAsnLys 180
Db |||||||
QY 614 -----AATACATAATATCAATTAAGTTGCTCTAAAGATTGCAA 652
Db |||||||
QY 181 AsnThrHisAsnIleAsnLysAsnThrHisAsnIleAsnLysLeuSerLysGluLeuGln 200
Db |||||||
QY 653 ACTGTTTGTAGCCAACTAGCATCTGCTATGTTAGTGAACCAAAATGTTGTAGGCAAA 712
Db |||||||
QY 201 ThrGlyLeuAlaAsnGlnSerAlaLeuSerMetLeuValGlnProAsnGlyValGlyLys 220
Db |||||||
QY 713 ACGAGCGTTTCTGCTGCGGTAGGAGTTATAGAGATAAAACTGCATTAGCCATTGGTGTG 772
Db |||||||
QY 221 ThrSerValSerAlaAlaValGlyTyrArgAspLysThrAlaLeuAlaIleGlyVal 240
Db |||||||
QY 773 GGCTCAGCATTTACTGATCGTTTACCGCTTAAGCGGGGTAGCGTTCAATACCTACAAT 832
Db |||||||
QY 241 GlySerArgIleThrAspArgPheThrAlaLysAlaGlyValAlaPheAsnThrTyrAsn 260
Db |||||||
QY 833 GCGGCGCATGTTATGTTGCTTCTGTTGGTTGATGAATTC 871
Db |||||||
QY 261 GlyGlyMetSerTyrGlyAlaSerValGlyTyrGluPhe 273
Db |||||||
RESULT 8
ID AAB31708 standard; protein; 242 AA.
XX
AC AAB31708;
XX
30-APR-2001 (first entry)
XX
Amino acid sequence of the DsrA protein from strain CIP542 (Can).
XX
DsrA protein; outer membrane protein; serum resistance; vaccine;
XX
chancroid disease; genital ulcer disease.
XX
Haemophilus ducreyi.
XX
W0200104138-A1.
PN
```

```
XX 18-JAN-2001.
PD
XX 07-JUL-2000; 2000WO-US018834.
PF
XX 09-JUL-1999; 99US-0143257P.
PR
XX (UYNC-) UNIV NORTH CAROLINA.
PA (ELKI/) ELKINS C.
XX
XX Elkins C;
XX
XX WPI; 2001-138311/14.
DR N-PSDB; AAF25265.
XX
XX Novel purified DsrA protein of Haemophilus ducreyi which confers serum
PT resistance to the bacteria used to produce vaccines that induce immune
PT response against the bacteria in subject at risk of developing chancroid.
XX
XX Claim 7; Page 55; 80pp; English.
XX
CC The present sequence represents a DsrA 30 kDa protein of Haemophilus
CC ducreyi. DsrA is an outer membrane protein that confers serum resistance
CC to the bacteria. DsrA antisense oligonucleotides are useful for detecting
CC a polynucleotide which encodes DsrA in a biological sample. The DsrA
CC polypeptide is used to produce vaccine compositions, which are useful for
CC inducing a protective immune response in a subject at risk of developing
CC Haemophilus ducreyi infection e.g. chancroid, a genital ulcer disease
CC transmitted by sexual contact. DsrA, its catalytic or immunogenic
CC fragments or peptides can be used for screening libraries of compounds in
CC a variety of drug screening techniques. The proteins and peptides may be
CC used as antigens in immunoassays for detection of Haemophilus ducreyi in
CC various tissues and body fluids, e.g. blood, spinal fluid, sputum etc. .
CC
XX The nucleic acids are useful for the preparation of DsrA proteins
SQ Sequence 242 AA;
Alignment Scores:
Pred. No.: 3,7e-138 Length: 242
Score: 1259.00 Matches: 242
Percent Similarity: 100.00% Conservativity: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 61.47% Indels: 0
DB: 4 Gaps: 0
US-10-030-529A-1 (1-1168) x AAB31708 (1-242)
QY 101 ATCAAAATTAATGTTTGTAGTTGCGTAGTGGATTAGCTTGTCTTACTATTACAAATG 160
Db |||||||
QY 1 MetLysIleLysCysLeuValAlaValAlaValGlyLeuAlaCysSerThrIleThrMet 20
Db |||||||
QY 161 GCTCAGCAGCGCCCAAAAGTTTGTGAGTATCTTTTGTATAGCTATGAGTATGACTAT 220
Db |||||||
QY 21 AlaGlnGlnProProLysPheAlaGlyValSerSerLeuTyrSerTyrGluTyrAspTyr 40
Db |||||||
QY 221 GGTAAAGGTTAAATGGACTGTCTTAATGAAGCGCGTTTCGATATTAAGTCCAGGATT 280
Db |||||||
QY 41 GlyLysGlyLysTrpThrTrpSerAsnGluGlyGlyPheAspIleLysValProGlyIle 60
Db |||||||
QY 281 AAAATGAAGCCAAAGAAATGATTTCTTAACAGCTACTTATCTTGAATTACAGCATAT 340
Db |||||||
QY 61 LysMetLysProLysGluTrpIleSerLysGlnAlaThrTyrLeuGluLeuGlnHisTyr 80
Db |||||||
QY 341 ATGCTTATATCTCTGTTCTGTCGACATATGCTCTCGCGGTTTCTCTAGCCTTACTG 400
Db |||||||
QY 81 MetProTyrThrProValLeuValThrTyrAlaProGlyValSerProSerProIleLeu 100
Db |||||||
QY 401 TTATATCCGATGTTCTGATCCTGATCAACTGGAAATAAATCGCAGAGCTGAAATGAAAT 460
Db |||||||
QY 101 LeuTyrProMetSerAspProAspGlnLeuGlyIleAsnArgGlnGlnLeuLysLeuAsn 120
Db |||||||
QY 461 TTGTATAGTTATTTTAAACGATTTTAAAGTAAATTAAGTCTTGTGACCGT 520
Db |||||||
```

Db 121 LeuTyrSerTyrPheAsnAspLeuArgHisAspPheLysLeuLysValLeuAspAlaArg 140
Qy 521 ATTTCCAAAATAAACAATAATTGATCTATAAGTAAATATTTACTAGAACTGGGTACT 580
Db 141 ILeSerLysAsnLysGlnAsnIleAspThrIleSerLysTyrLeuLeuGluGlyThr 160
Qy 581 TATTAGATGATCTTATCGTATGATGGAACAAATACACATAATATCAATAGTTGTCT 640
Db 161 TyrLeuAspAspSerTyrArgMetMetGluGlnAsnThrHisAsnIleAsnLysLeuSer 180
Qy 641 AAGAATGTGCAACTGTTAGCCCAACCAATCAGCATGCTATGTTAGTGCACCAAT 700
Db 181 LysGluLeuGlnThrGlyLeuAlaAsnGlnSerAlaLeuSerMetLeuValGlnProAsn 200
Qy 701 GGTGTAGCAAAACGAGGTTTCTGCTCGCTAGGAGTTATAGAGATAAACTGCATTA 760
Db 201 GlyValGlyLysThrSerValSerAlaAlaValGlyTyrArgAspLysThrAlaLeu 220
Qy 761 GCCATTGGTGTGGCTCACGATTACTGATCGCTTTACCGCTAAAGCGGGGTAGCGTTC 820
Db 221 AlaIleGlyValGlySerArgIleThrAspArgPheThrAlaLysAlaGlyValAlaPhe 240
Qy 821 AATACC 826
Db 241 AsnThr 242
RESULT 9
AAB31710
ID AAB31710 standard; protein; 263 AA.
AC AAB31710;
XX
XX 30-APR-2001 (first entry)
DT
XX Amino acid sequence of the DsrA protein from strain CH1A.
DE
XX
XX DsrA protein; outer membrane protein; serum resistance; vaccine;
KW chancroid disease; genital ulcer disease.
KW
XX Haemophilus ducreyi.
OS
XX WO200104138-A1.
PN
XX 18-JAN-2001.
PD
XX 07-JUL-2000; 2000WO-US018834.
PF
XX 09-JUL-1999; 99US-0143257P.
PR
XX (UYNCL) UNIV NORTH CAROLINA.
PA
XX (ELKI/) ELKINS C.
PI
XX Elkins C;
XX
XX WPI; 2001-138311/14.
DR
XX N-PSDB; AAF25267.
XX
PT Novel purified DsrA protein of Haemophilus ducreyi which confers serum
PT resistance to the bacteria used to produce vaccines that induce immune
PT response against the bacteria in subject at risk of developing chancroid.
XX
PS Claim 7; Page 56; 80pp; English.
XX

CC The present sequence represents a DsrA 30 kDa protein of Haemophilus
CC ducreyi. DsrA is an outer membrane protein that confers serum resistance
CC to the bacteria. DsrA antisense oligonucleotides are useful for detecting
CC a polynucleotide which encodes DsrA in a biological sample. The DsrA
CC polypeptide is used to produce vaccine compositions, which are useful for
CC inducing a protective immune response in a subject at risk of developing
CC Haemophilus ducreyi infection e.g. chancroid, a genital ulcer disease
CC transmitted by sexual contact. DsrA, its catalytic or immunogenic
CC fragments or peptides can be used for screening libraries of compounds in
CC a variety of drug screening techniques. The proteins and peptides may be

CC used as antigens in immunoassays for detection of Haemophilus ducreyi in
CC various tissues and body fluids, e.g. blood, spinal fluid, sputum etc. .
CC The nucleic acids are useful for the preparation of DsrA proteins
XX
SQ Sequence 263 AA;
Alignment Scores:
Pred. No.: 8,22e-132 Length: 263
Score: 1205.00 Matches: 235
Percent Similarity: 92.05% Conservative: 8
Best Local Similarity: 89.02% Mismatches: 13
Query Match: 58.84% Indels: 8
DB: Gaps: 2

US-10-030-529A-1 (1-1168) x AAB31710 (1-263)
Qy 101 ATCAAAATTAATGTTTGTAGTTGCGGTAGTGGGATTAGCTTGTCTACTATTACAAATG 160
Db 1 MetLysIleLysCysLeuValAlaValGlyLeuAlaCysSerThrIleThrMet 20
Qy 161 GCTCAGCAGCGCCAAAGTTTGTGGAGTATCTTCTTTGTATAGTATGATGACTAT 220
Db 21 AlaGlnGlnProProlLysPheAlaGlyValSerSerLeuAspSerTyrGluTyrAspTyr 40
Qy 221 GGTAAAGGTAAATCGACTTCGTCTAATGAAGCGGTTTCGATATATAAGTCCAGGATT 280
Db 41 GlyLysGlyLysTrpThrTrpSerGluLysAspGlyPheAspIleLysAlaProGlyIle 60
Qy 281 AAAATGAGCCAAAGNATGGATTCTTAAACAGGCTACTTATCTTGAATTACAGCATAT 340
Db 61 LysMetLysProlLysLysTrpIleSerArgGlnAlaThrTyrLeuGlyLeuGlnHisTyr 80
Qy 341 ATGCTTATATCTCTGTTCTCGTGACATATGCTCGCGTCTTCTCTAGCCCTATATCTG 400
Db 81 MetProTyrThrProValLeuValThrTyrAla---SerAlaGluProAsnThrValLeu 99
Qy 401 TTATATCCGATGTCTGATCTGATCAACTTGGATAAATAATCGGACGAGCTGAATGAAT 460
Db 100 LeuTyrProMetProAspProAspGlnLeuGlyLeuAsnArgGlnGlnLeuLysLeuAsn 119
Qy 461 TTGTATAGTTATTTTAACGATTTAAGACACGATTTTAAATTTAAAGTCTTGTATGCAGT 520
Db 120 LeuTyrSerTyrPheAsnAspLeuArgHisGlyPheLysLeuAsnValLeuAspAlaArg 139
Qy 521 ATTTCCAAAATAAACAATAATTGATCTATAAGTAAATATTTACTAGAACTGGGTACT 580
Db 140 ILeSerGlnAsnLysGlnAsnIleAspThrIleSerGluTyrLeuLeuLysLeuGlyThr 159
Qy 581 TATTAGATGATCTTATCGTATGATGGAACAA-----AATACA 619
Db 160 TyrLeuAspSerSerTyrArgMetMetGluGlnAsnThrHisAsnIleAsnLysAsnThr 179
Qy 620 CATATATCAATAAGTTGTCTTAAAGATTCGAAATCGTTTAGCCCAACCATCAGCATG 679
Db 180 HisAsnIleAsnLysLeuSerLysGluLeuGlnThrGlyLeuAlaAsnGlnSerAlaLeu 199
Qy 680 TCTATGTTAGTCAACCAAAATGGTGTAGGCAAAACGAGCGTTTCTGCTCGGTAGGAGGT 739
Db 200 SerMetLeuValGlnProAsnGlyValGlyLysThrSerValSerAlaAlaValGlyGly 219
Qy 740 TATAGATATAAACTGCATTTAGCCATTGGTGTGCGCTCAGCATTTACTGATTCGCTTACC 799
Db 220 TyrArgAspLysThrAlaLeuAlaIleGlyValGlySerArgIleThrAspArgPheThr 239
Qy 800 GCTTAACCGGTGTAGGTTTCAATACCTACATGCGCGCATGCTTATGGTCTCTGTT 859
Db 240 AlaLysAlaGlyValAlaPheAsnThrTyrAsnGlyGlyMetSerTyrGlyAlaSerVal 259
Qy 860 GGTATGAATTC 871
Db 260 GlyTyrGluPhe 263

RESULT 10

ABU07927

ID ABU07927 standard; protein; 323 AA.

XX AC ABU07927;

XX DT 23-MAY-2003 (first entry)

XX DE Neisserial adhesin A (NadA) allele C.

XX KW Neisserial adhesin A; NadA; antibacterial; immunostimulant; vaccine;

KW neisserial infection; meningitis; bacterial meningitis; bacteraemia;

KW systemic immunity; mucosal immunity; allele.

XX OS Neisseria meningitidis.

XX PN WO2003010194-A2.

PX PD 06-FEB-2003.

PF 26-JUL-2002; 2002WO-IB003396.

XX 27-JUL-2001; 2001GB-00018401.

PR 06-SEP-2001; 2001GB-00021591.

PR 14-MAY-2002; 2002GB-00011025.

XX (CHIR-) CHIRON SPA.

PA Arico M, Comanducci M;

PI WPI; 2003-248057/24.

PX N-PSDB; ABX933390.

DR New Neisserial adhesin A protein and nucleic acids, useful for preventing or treating meningitis, particularly bacterial meningitis, and bacteremia, and for eliciting an systemic and/or mucosal immunity.

PT Claim 1; Page 76; 79pp; English.

PS The invention describes a Neisserial adhesin (NadA) comprising a 362, 398, 405, 364, 407, 391, 393, 405, 107, 355, 357, 323, or 319 residue amino acid sequence given in the specification, or an amino acid sequence having at least 50 % identity to the amino acid sequences, or a fragment of them. The NadA protein, or nucleic acid encoding NadA protein is useful in the manufacture of a medicament for preventing Neisserial infection in a mammal, such as an infection of Neisseria meningitidis from hypervirulent lineages ET-5, EY-37 and cluster M4. The NadA protein is useful for preventing or treating diseases, specifically meningitis (particularly bacterial meningitis) and bacteraemia, and for eliciting an systemic and/or mucosal immunity. This is the amino acid sequence of neisserial adhesin A (NadA) allele C

XX Sequence 323 AA;

SQ Alignment Scores:

Pred. No.:	4.85e-16	Length:	323
Score:	218.00	Matches:	75
Percent Similarity:	43.18%	Conservative:	39
Best Local Similarity:	28.41%	Mismatches:	94
Query Match:	10.64%	Indels:	56
DB:	6	Gaps:	8

US-10-030-529A-1 (1-1168) x ABU07927 (1-323)

Qy	83	TACTTAATGAGGTGATTGAATAAATTAATTTAGTTCGGTAGTGGATTAGCTTG	142
Db	115	ValValAsnAspIleSerAlaAspValLysAlaAsnThrAlaAlaIleGlyGluAsnLys	134
Qy	143	TCTACTATTACAACAAATGCTCAGCAGCGCCCAAGTTTGCCTGGAGTATCTTCTTGAT	202
Db	135	AlaAlaIleAlaThrLysAlaAspLysThr-----	144
Qy	203	AGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	262

Db 145 ----- GluLeuAspLysValSerGlyLysValThr----- GluAsnGluThrAla----- 155

Qy 263 ATTAAGAATGCCAGGATTAAAAATGAAGCCAAAAGAATGAGATTTCTTAACAGCGCTACTTAT 322

Db 160 ----|----|----|----|----|----|----|----|----|----|----|----|----| 166

Qy 323 CTGTGAATTACAGCATTATATGCTTATATCTCTCTCTGTCGATGACATATGCTCTCTGGCGGT 382

Db 167 AlaAspVal-----TyrThrLysAlaGluVal----- 175

Qy 383 TCTCTAGCCCTATATCTGTATATCCGATGTCGATCTCTGATCACTTGGAAATAAACCG 442

Db 176 -----TyrThrLysGlnGluSerAspAsnArgPheValLysile 188

Qy 443 CAGCAGCTGAAATGAAATTTGATATAGTTATTTTACGATTTTAAAGACAGCATTTTAAATTA 502

Db 189 SerAspGlyIleGlyAsnLeuAsnThrThrAlaAsnGlyLeu----- GluThrArgLeu 206

Qy 503 AAAGTCTTGATGCACGATTTTCCAAAAATAAAATAATTTGATCTATAAGTAATAATAT 562

Db 207 AlaAlaAlaGluGlnSerValAlaAspHisGlyThrArgLeuAlaSerAlaGluLysSer 226

Qy 563 TTACTAGAACTGGGTACTTATTAGATGATTTCTTATCGTATGATGGAACAAATACACAT 622

Db 227 IleThrGluHisGlyThrArgLeuAsnGlyLeuAspArg----- 239

Qy 623 AATATCAATTAAGTTCTTAAGAATTTGCAAATGTTAGTGGTTCAGCAACCAATCAGCATTTGCT 682

Db 240 ThrValSerAspLeuArgLysGluThrArgGlnGlyLeuAlaGluGlnAlaLeuSer 259

Qy 683 ATGTAGTCACCAATGGTGTAGCAGCAAAACCAGCGTTTTCTGCTGCGGTAGGAGTTAT 742

Db 260 GlyLeuPheGlnProTyrAsnValGlyArgPheAsnValThrAlaAlaValGlyGlyTyr 279

Qy 743 AGAGATAAACTGCATTAGCCCATTCGTCGGCTCACGCATTACTGATCGCTTTACCGCT 802

Db 280 LysSerGluSerAlaValAlaIleGlyThrGlyPheArgPheThrGluAsnPheAla 299

Qy 803 AAAGCGGGTGTAGCGTTCAATACCTACAAATGCG---GCGATGTCTTTATGGTCTTCTGTT 859

Db 300 LysAlaGlyValAlaValGlyThrSerGlySerSerAlaAlaTyrHisValGlyVal 319

Qy 860 GGTATTGAATTC 871

Db 320 AsnTyrGluTrp 323

RESULT 11

ABU07925

ID ABU07925 standard; protein; 355 AA.

XX AC ABU07925;

XX XX

DT 23-MAY-2003 (first entry)

XX DE Neisserial adhesin A (NadA) allele 1/2 chimera (strain 95330).

XX KW Neisserial adhesin A; NadA; antibacterial; immunostimulant; vaccine;

KW neisserial infection; meningitis; bacterial meningitis; bacteraemia;

KW systemic immunity; mucosal immunity; allele; strain 95330; chimeric.

XX OS Neisseria meningitidis.

XX OS Synthetic.

XX PN WO2003010194-A2.

PX PD 06-FEB-2003.

PF 26-JUL-2002; 2002WO-IB003396.

XX 27-JUL-2001; 2001GB-00018401.

PR 06-SEP-2001; 2001GB-00021591.

PR 14-MAY-2002; 2002GB-00011025.

XX

```
PA (CHIR-) CHIRON SPA.
XX
XX
PI Arico M, Comanducci M;
XX
XX WPI; 2003-248057/24.
XX
XX New Neisserial adhesin A protein and nucleic acids, useful for preventing
PT or treating meningitis, particularly bacterial meningitis, and
PT bacteremia, and for eliciting an systemic and/or mucosal immunity.
XX
XX Claim 1; Page 76; 79pp; English.
XX
XX The invention describes a Neisserial adhesin (NadA) comprising a 362,
CC 398, 405, 364, 400, 407, 391, 393, 405, 107, 355, 357, 323, or 319
CC residue amino acid sequence given in the specification, or an amino acid
CC sequence having at least 50 % identity to the amino acid sequences, or a
CC fragment of them. The NadA protein, or nucleic acid encoding NadA protein
CC is useful in the manufacture of a medicament for preventing Neisserial
CC infection in a mammal, such as an infection of Neisseria meningitidis
CC from hypervirulent lineages ET-5, EV-37 and cluster A4. The NadA protein
CC is useful for preventing or treating diseases, specifically meningitis
CC (particularly bacterial meningitis) and bacteraemia, and for eliciting an
CC systemic and/or mucosal immunity. This is the amino acid sequence of a
CC neisserial adhesin A (NadA) allele 1/2 chimera (strain 95330)
XX
SQ Sequence 355 AA;

Alignment Scores:
Pred. No.: 6.61e-16 Length: 355
Score: 217.00 Matches: 52
Percent Similarity: 53.79% Conservative: 19
Best Local Similarity: 39.39% Mismatches: 37
Query Match: 10.60% Indels: 24
DB: 6 Gaps: 3

US-10-030-529A-1 (1-1168) x ABU07925 (1-355)
QY 479 GATTTAAGACACGAGTATTTAAATTAAGTCTTGATGCACGTATTTCCAAAAATAACAA 538
DB 247 AspIleLysAlaAspIleAlaThrAsnLysAlaAsp-----IleAlaLysAsnSerAla 264
QY 539 AATATTGATACATAAGTAATAATTTACTAGAACTGGGTACTTATTTAGATGATCTTAT 598
DB 265 ArgIleAspSerLeuAspLys----- 271
QY 599 CGTATGATGGAACAAATAACATATATCAATAGTTGCTTAAGAAATTCGAACCTGGT 658
DB 272 -----AsnValAlaAsnLeuArgLysGluThrArgGlnGly 283
QY 659 TTAGCCAAACCAATCAGCATTTGCTATGTAGTGCACCAAAATGGTGTAGGCAAAACGAGC 718
DB 284 LeuAlaGluGlnAlaAlaLeuSerGlyLeuPheGlnProTyrAsnValGlyArgPheAsn 303
QY 719 GTTTCTGTCGGGTAGGAGTTATAGAGATAAAACTGCATTAGCCATTGGTGTGCGGTCA 778
DB 304 ValThrAlaAlaValGlyGlyTyrTyrSerGluSerAlaValAlaIleGlyThrGlyPhe 323
QY 779 GCATTACTGATCGCTTTACCGCTTAACCGGGTGTAGCGTTCAATACCTACATGGC--- 835
DB 324 ArgPheThrGluAsnPheAlaAlaLysAlaGlyValAlaValGlyThrSerSerGlySer 343
QY 836 GGCATGTCCTTGGTGGCTCTCTGGTTATGAATTC 871
DB 344 SerAlaAlaTyrHisValGlyValAsnTyrGluIrrp 355

RESULT 12
ABU07926
ID ABU07926 standard; protein; 357 AA.
XX
XX AC ABU07926;
XX
XX DT 23-MAY-2003 (first entry)
XX
```

```
DE Neisserial adhesin A allele 1/2 chimera (first ATG) in strain 95330.
XX
XX Neisserial adhesin A; NadA; antibacterial; immunostimulant; vaccine;
XX neisserial infection; meningitis; bacterial meningitis; bacteraemia;
XX systemic immunity; mucosal immunity; allele; strain 95330; chimeric.
XX
XX Neisseria meningitidis.
OS Synthetic.
XX
XX W02003010194-A2.
XX
XX 06-FEB-2003.
XX
XX 26-JUL-2002; 2002WO-IB003396.
XX
XX 27-JUL-2001; 2001GB-00018401.
PR 06-SEP-2001; 2001GB-00021591.
PR 14-MAY-2002; 2002GB-00011025.
XX
XX (CHIR-) CHIRON SPA.
XX
XX Arico M, Comanducci M;
XX
XX WPI; 2003-248057/24.
XX
XX New Neisserial adhesin A protein and nucleic acids, useful for preventing
PT or treating meningitis, particularly bacterial meningitis, and
PT bacteremia, and for eliciting an systemic and/or mucosal immunity.
XX
XX Claim 1; Page 76; 79pp; English.
XX
XX The invention describes a Neisserial adhesin (NadA) comprising a 362,
CC 398, 405, 364, 400, 407, 391, 393, 405, 107, 355, 357, 323, or 319
CC residue amino acid sequence given in the specification, or an amino acid
CC sequence having at least 50 % identity to the amino acid sequences, or a
CC fragment of them. The NadA protein, or nucleic acid encoding NadA protein
CC is useful in the manufacture of a medicament for preventing Neisserial
CC infection in a mammal, such as an infection of Neisseria meningitidis
CC from hypervirulent lineages ET-5, EV-37 and cluster A4. The NadA protein
CC is useful for preventing or treating diseases, specifically meningitis
CC (particularly bacterial meningitis) and bacteraemia, and for eliciting an
CC systemic and/or mucosal immunity. This is the amino acid sequence of a
CC neisserial adhesin A (NadA) allele 1/2 (first ATG start) chimera (strain
CC 95330)
XX
XX SQ Sequence 357 AA;

Alignment Scores:
Pred. No.: 6.63e-16 Length: 357
Score: 217.00 Matches: 52
Percent Similarity: 53.79% Conservative: 19
Best Local Similarity: 39.39% Mismatches: 37
Query Match: 10.60% Indels: 24
DB: 6 Gaps: 3

US-10-030-529A-1 (1-1168) x ABU07926 (1-357)
QY 479 GATTTAAGACACGAGTATTTAAATTAAGTCTTGATGCACGTATTTCCAAAAATAACAA 538
DB 249 AspIleLysAlaAspIleAlaThrAsnLysAlaAsp-----IleAlaLysAsnSerAla 266
QY 539 AATATTGATACATAAGTAATAATTTACTAGAACTGGGTACTTATTTAGATGATCTTAT 598
DB 267 ArgIleAspSerLeuAspLys----- 273
QY 599 CGTATGATGGAACAAATAACATATATCAATAGTTGCTTAAGAAATTCGAACCTGGT 658
DB 274 -----AsnValAlaAsnLeuArgLysGluThrArgGlnGly 285
QY 659 TTAGCCAAACCAATCAGCATTTGCTATGTAGTGCACCAAAATGGTGTAGGCAAAACGAGC 718
DB 286 LeuAlaGluGlnAlaAlaLeuSerGlyLeuPheGlnProTyrAsnValGlyArgPheAsn 305
```


Job time : 69 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: May 13, 2004, 08:18:07 ; Search time 17.5 seconds

(without alignments)

6950.622 Million cell updates/sec

Title: US-10-030-529a-1

Perfect score: 2048

Sequence: 1 ataaatcgcattgacatt.....aagccgtaaggtggcgac 1168

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 283362

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlp
-Q=/cgn2/1/USPTO_epool_p/US10030529/runat_13052004_081452_2600/app_query.fasta_1.1351
-DB=SwissProt 42 -QFMT=fastan -SUFFIX=rs -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -NATRX=bloum62 -TRANS=human40.cdi -LIST=45
-LOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTPMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10030529@cgn 1_16 @runat_13052004_081452_2600 -NCPU=6 -ICPU=3
-NO MMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	186.5	9.1	455	1 YADA_YEREN	P31489 yersinia en
2	182	8.9	434	1 YADA_YEREN	P10858 yersinia ps
c 3	144	7.2	420	1 YFJD_HAEIN	Q57017 haemophilus
c 4	130	6.5	428	1 YFJD_ECOLI	P37908 escherichia
5	102.5	5.0	498	1 FLID_BACSU	P39738 bacillus su
6	101	4.9	1690	1 KF1A_HUMAN	Q12756 homo sapien
7	99.5	4.9	831	1 P1A1_SYNY3	P74397 synechocyst
8	99.5	4.9	1695	1 KF1A_MOUSE	P33173 mus musculu
9	99	4.8	795	1 LON_MYCGE	P47481 mycoplasma
10	95	4.6	2214	1 POLG_CXA24	P36290 c genome po
11	94.5	4.6	1036	1 Y946_ARCFU	Q29316 archaeoglob
12	92.5	4.5	588	1 CMC2_CABEL	Q20799 caenorhabdi
13	92.5	4.5	727	1 MFP1_ARATH	Q91W85 arabidopsis
14	92.5	4.5	824	1 HELI_HSV6U	P52356 human herpe
15	92	4.5	444	1 CHLB_CHLPT	P37824 chlamydomon
16	90.5	4.4	290	1 CYOA_BUCAP	Q8K993 buchnera ap
17	90.5	4.4	819	1 LON_CHLPN	P929f4 chlamydia p
18	90.5	4.4	1226	1 DSRA_HUMAN	P55265 homo sapien

19	90.5	4.4	2716	1 OSA_DROME	Q8in94 drosophila
20	90	4.4	576	1 FTSI_BUCBP	Q89aq0 buchnera ap
21	90	4.4	758	1 PXA1_YEAST	P41909 saccharomyc
22	90	4.4	1328	1 POLA_TOBAC	P10978 nicotiana t
23	89.5	4.4	824	1 HELI_HSV6Z	P52450 human herpe
24	89.5	4.4	956	1 UVRA_LISIN	Q928a5 listeria in
25	88.5	4.3	4543	1 LRPI_CHICK	P98157 gallus gall
26	88	4.3	2206	1 POLG_POL1M	P03299 p genome po
27	87.5	4.3	643	1 TBP4_EPTST	Q90501 eptaretus
28	87.5	4.3	820	1 HELI_HSV7J	P52357 human herpe
29	87	4.2	806	1 LON_BORBU	Q59185 borrelia bu
30	87	4.2	2208	1 POLH_POL1M	P03300 p genome po
31	86.5	4.2	524	1 CLPP_CHLRE	P42380 chlamydomon
32	86	4.2	473	1 Y085_UREPA	Q9pr60 ureaplasma
33	86	4.2	677	1 YD64_MYCPN	P75417 mycoplasma
34	86	4.2	766	1 TKL1_MOUSE	Q8C0V0 mus musculu
35	86	4.2	920	1 YML1_YEAST	Q33758 saccharomyc
36	85.5	4.2	218	1 GTH5_ARATH	P42769 arabidopsis
37	85.5	4.2	419	1 ENO_SULSO	Q972j3 sulfolobus
38	85.5	4.2	522	1 LEUL_SHEON	Q8egn2 shewanella
39	85	4.2	810	1 HPUB_NEIMC	P96949 neisseria m
40	85	4.2	2209	1 POLG_POL1S	P03301 p genome po
41	84.5	4.1	491	1 G6PD_BUCAI	P57405 buchnera ap
42	84.5	4.1	947	1 LKTA_PASSP	P55123 pasteurella
43	84.5	4.1	1790	1 VIT_ANTGR	Q05808 anthomomus
44	84	4.1	438	1 GLGA_TRECA	P58395 thermus cal
45	84	4.1	662	1 YME1_SCHMA	P46508 schistosoma

ALIGNMENTS

RESULT 1

YADA_YEREN					
ID	YADA_YEREN	STANDARD;	PRT;	455 AA.	
AC	P31489;				
DT	01-JUL-1993 (Rel. 26, Created)				
DT	01-JUL-1993 (Rel. 26, Last sequence update)				
DT	01-OCT-1996 (Rel. 34, Last annotation update)				
DE	Invasin precursor (Outer membrane adhesin).				
GN	YADA OR YOPA OR INVA OR YOPI.				
OS	Yersinia enterocolitica.				
OG	Plasmid pYV.				
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;				
OC	Enterobacteriaceae; Yersinia.				
OX	NCBI_TaxID=630;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=6471/76 / Serotype O:3;				
RX	MEDLINE=95020586; PubMed=7934875;				
RA	Tamm A., Tarkkanen A., Korhonen T.K., Kuusela P., Toivanen P.,				
RA	Skurnik M.;				
RT	"Hydrophobic domains affect the collagen-binding specificity and				
RT	surface polymerization as well as the virulence potential of the Yada				
RT	protein of Yersinia enterocolitica.";				
RL	Mol. Microbiol. 10:995-1011(1993).				
CC	-1- FUNCTION: INVASIN IS A PROTEIN THAT ALLOWS ENTERIC BACTERIA TO				
CC	PENETRATE CULTURED MAMMALIAN CELLS. THE ENTRY OF INVASIN IN THE				
CC	CELL IS MEDIATED BY BINDING SEVERAL BETA-1 CHAIN INTEGRINS. THIS				
CC	PROTEIN IS AN ADHESIN FORMING A FIBRILLAR MATRIX ON THE CELL				
CC	SURFACE.				
CC	-1- SUBCELLULAR LOCATION: Outer membrane.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/				
CC	or send an email to license@isb-sib.ch).				
CC	-----				
DR	EMBL; X13882; CAA32086.1; -				
DR	PIR; S04912; S04912.				
DR	InterPro; IPR008126; Adhesion.				

[illegible]

Percent Similarity:	33.57%	Conservative:	52
Best Local Similarity:	22.98%	Mismatches:	92
Query Match:	4.93%	Indels:	50
DB:	1	Gaps:	11
US-10-030-529A-1 (1-1168) x KP1A_HUMAN (1-1690)			
Qy	65	ATCAAGATTGACAATATTTTACTTAAATGAGGTGATTAATGAAAAATTAATGTTAGTTGCC	124
	:::>:::::	:::>:::::	:::>:::
Db	10	ValargValargProPheAsnSerArgGluMetSerArgSerpSerLysCysIleIleGln	29
Qy	125	GTAGTGGGATTAGTGTGTTCTACTATTATCAACAATGGCTGACGACGCCCAAGTTTGTCT	184

Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

NCBI_TaxID=9606;

(1)

SEQUENCE FROM N.A.

TISSUE=Brain;

MEDLINE=96299637; PubMed=8661001;

Furlong R.A., Zhou C.Y., Ferguson-Smith M.A., Affara N.A.;

"Characterization of a kinesin-related gene AFSV, within the tuberous

sclerosis locus (TSC1) candidate region on chromosome 9q34.,";

Genomics 33:421-429(1996).

-!- FUNCTION: Motor for anterograde axonal transport of synaptic

vesicle precursors (By similarity).

-!- SUBUNIT: Monomer (By similarity).

-!- SIMILARITY: Belongs to the kinesin-like protein family. Unc-104

subfamily.

-!- SIMILARITY: Contains 1 FHA domain.

-!- SIMILARITY: Contains 1 PH domain.

This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>

or send an email to license@isb-sib.ch).

Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
NCBI TaxID=9606;

Db 167 ArgValArgGluHisProLeu---LeuglyProTyrValGluAspLeuSerLysLeuAla 185
QY 608 GAACAAATACACATAATATCAATAAGTTG 637
Db 186 ValThrSerTyrAsnAspIleGlnAspLeu 195
RESULT 9
LON_MYCGE
ID LON_MYCGE STANDARD; PRT; 795 AA.
AC P47481; Q49276;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE ATP-dependent protease La (EC 3.4.21.53).
GN LON OR MG239.
OS Mycoplasma genitalium.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2097;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33530 / G-37;
RX MEDLINE=96026346; PubMed=7569993;
RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
RA Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,
RA Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhrmann J.L.,
RA Nguyen D.T., Utterback T.R., Saudek D.M., Phillips C.A., Merrick J.M.,
RA Tomb J.-F., Dougherty B.A., Brott K.F., Hu P.-C., Lucier T.S.,
RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.,
RT "The minimal gene complement of Mycoplasma genitalium."
RL Science 270:397-403(1995).
RN [2]
RP SEQUENCE OF 484-606 FROM N.A.
RC STRAIN=ATCC 33530 / G-37;
RX MEDLINE=94075230; PubMed=8253680;
RA Peterson S.N., Hu P.-C., Bott K.F., Hutchison C.A. III;
RT "A survey of the Mycoplasma genitalium genome by using random
RT sequencing."
RL J. Bacteriol. 175:7918-7930(1993).
CC -!- FUNCTION: DEGRADATES SHORT-LIVED REGULATORY AND ABNORMAL PROTEINS
CC IN PRESENCE OF ATP. DEGRADATES THE REGULATORY PROTEINS RCSA AND
CC SULA. HYDROLYZES TWO ATPs FOR EACH PEPTIDE BOND CLEAVED IN THE
CC PROTEIN SUBSTRATE (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: Hydrolysis of large proteins such as globin,
CC casein and denaturated serum albumin, in presence of ATP.
CC -!- SUBUNIT: Homotetramer (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to peptidase family S16.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb.ch/announce/>
CC or send an email to licences@isb-sib.ch).
CC
CC EMBL; U39703; AAC71460.1; -.
CC EMBL; U02148; AAD12428.1; -.
CC PIR; D64226; D64226.
CC MEROPS; S16.004; -.
CC TIGR; MG239; -.
CC InterPro; IPR003593; AAA_ATPase.
CC InterPro; IPR003959; AAA_ATPase_cent.
CC InterPro; IPR008269; Pept_S16_C.
CC InterPro; IPR004815; Pept_S16_N.
CC InterPro; IPR003111; Pept_S16_N.
CC InterPro; IPR008268; Peptid_S16_AS.
CC InterPro; IPR001984; peptidase_S16.
CC Pfam; PF00004; AAA; 1.
CC Pfam; PF02190; LON; 1.
CC Pfam; PF05362; Lon_C; 1.
CC PRINTS; PR00830; ENDOLAPTASE.
CC SMART; SM00382; AAA; 1.

DR SMART; SM00464; LON; 1.
DR TIGRFAMS; TIGR00763; Lon; 1.
DR PROSITE; PS01046; LON_SER; 1.
KW Hydrolase; Serine protease; ATP-binding; Complete proteome.
FT DOMAIN 304 308 POLY-SER.
FT NP_BIND 379 386 ATP (POTENTIAL).
FT ACT_SITE 702 702 BY SIMILARITY.
FT CONFLICT 494 494 T -> S (IN REF. 2).
SQ SEQUENCE 795 AA; 89987 MW; 31C8188EA40F5E39 CRC64;

Alignment Scores:
Pred. No.: 0.655 Length: 795
Score: 99.00 Matches: 61
Percent Similarity: 38.46% Conservative: 44
Best Local Similarity: 22.34% Mismatches: 116
Query Match: 4.83% Indels: 52
DB: 1 Gaps: 10

US-10-030-529A-1 (1-1168) x LON_MYCGE (1-795)
QY 26 AATGTAAGGTAGATAAGAAAGTAATTTCTATATTTTACAAATCAAGATTGACAAATTATTGA 85
Db 163 AsnValProLysGlyLeuLysGlnLeuAspIleleThrPheLysLeuAlaAsnLeuVal 182
QY 86 CTTAATGAGGTGATTATCAAAATTAATTTAGTTGCGGTAGTGGGATTAGCTTGTCT 145
Db 183 ProAsnThrGluSerIleLysGlnAlaLeuGluGlu----- 195
QY 146 ACTATTACAAATGGCTGACGACGCCGCAAGTTT-----GCTGGAGTATCTCTTTG 199
Db 196 -----AsnGluIleAlaAsnArgLeuGluLysIleleGlnAlaGlyleGluAspLeu 213
QY 200 TATAGCTATGATGATGTAAGGTAAATGAGTCTGTCTTAATGAAGGCGGTTTC 259
Db 214 GlnLysleGln---AspTyrGlyArgSerLys-----AsnLysGluThrGluPhe 229
QY 260 GATATTAAAGTCCAGGATTAAATGAAGCAAAAGAAATGGATTCTTAAACAGGCTACT 319
Db 230 AspLysLeuAspSerLysIleThrArgLysIleAsnGluGlnLeuSerArgGln----- 247
QY 320 TATCTTGAATTACAGCATATATGCTTATATCTCTGTCTCTCGTGCATATGCTCCTGCG 379
Db 248 -----GlnArgAspPheTyrLeuArgGluLysLeuArgIleleleArgGluGluIleGly 265
QY 380 GTTTCCTCCTAGCCCTACTATGTTATATCCGATGCTGTGATCTCTGATCAACTTGGAAATAAT 439
Db 266 Ileser-----SerLysLysGluAspGluValAlaSerIle 277
QY 440 CGGCAGCAGCTGAAATTTGAATTTGTAT-----AGTTATTTTAAAGAT 481
Db 278 ArgLysLysLeuAspGluAsnProTyrProGluAlaIleLysLysArgIleLeuSerGlu 297
QY 482 TTAACACAGCATTTTAATTAATAAGTTCTTGATGACGCTATTTCACAAAATAACAAAT 541
Db 298 LeuGluHisTyrGluAsnSerSerSerSerGlnGluSerThrLeuThrLysThrTyr 317
QY 542 ATTGATACATAAGTAATAATTTTACTAGAACTG-----GGT 577
Db 318 IlespThr-----LeuLeuAsnLeuProTyrTrpGlnLysSerLysAspLeu 333
QY 578 ACTTATTTTAGATGATTTCTTATCGTATGATGAACAAATAACACATAATATCAATAGTTG 637
Db 334 SerAspValLysAsnLeuIleLysThrLeuAspLysAsnHisThrGlyLeuAspLysVal 353
QY 638 TCTAAGAAATTCGAAACTGGTTTATGAC-----AACCAATCAGCA 676
Db 354 LysGluArgIleValGluTyrLeuAlaValGlnLeuArgThrGlnLysAsnLysGlyPro 373
QY 677 TTGCTATGTTAGTGCACCAATATGCTAGCAAAACGAGCGTTCTCTCGCGGTAGGA 736
Db 374 IleMetCysLeuValGlyProGlyValGlyLysSerSerLeuAlaLysSerIleAla 393
QY 737 GGTTATAGAGATAAAACTGCATTAGCCATTGGTGTGCGC 775

```

Db 394 GluAlaLeuaspLysPheValLysIleSerLeugly 406
RESULT 10
POLG_CXA24
ID POLG_CXA24 STANDARD; PRT; 2214 AA.
AC P36290;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Genome polyprotein [Contains: Coat protein VP4 (P1A); Coat protein VP2
DE (P1B); Coat protein VP3 (P1C); Coat protein VP1 (P1D); Core protein
DE P2A; Core protein P2B; Core protein P2C; Core protein P3A; Genome-
DE linked protein VP3 (P3B); Picornain 3C (EC 3.4.22.28) (Protease 3C)
DE (P3C); RNA-directed RNA polymerase (EC 2.7.7.48) (P3D)].
OS Coxsackievirus A24 (strain EH24/70).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Enterovirus.
OX NCBI_TaxID=36404;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92271460; PubMed=1317075;
RA Supanaranond K., Takeda N., Yamazaki S.;
RT "The complete nucleotide sequence of a variant of Coxsackievirus A24,
RT an agent causing acute hemorrhagic conjunctivitis.";
RL Virus Genes 6:149-158(1992).
CC -!- FUNCTION: It is thought that the P2C protein attaches to vesicular
CC membranes and is associated with viral RNA synthesis.
CC -!- FUNCTION: P3C POLYPEPTIDE IS A PROTEASE THAT CLEAVES AT CERTAIN
CC O/G SITES IN THE POLYPROTEIN. IT MAY BE A CYSTEINE PROTEASE.
CC -!- CATALYTIC ACTIVITY: Selective cleavage of Gln-Gly bond in the
CC poliovirus polyprotein. In other picornavirus reactions Glu may be
CC substituted for Gln, and Ser or Thr for Gly.
CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC [RNA] (N).
CC -!- SUBUNIT: The virus capsid is composed of 60 icosahedral units,
CC each of which is composed of one copy each of proteins VP1, VP2,
CC VP3, and VP4.
CC -!- PTM: Specific enzymatic cleavages in vivo yield mature proteins.
CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D90457; -; NOT_ANNOTATED_CDS.
DR PIR; A48548; A48548.
DR HGSP; P03299; IAR7.
DR MEROPS; C03.001; -.
DR MEROPS; C03.020; -.
DR InterPro; IPR004004; Calici_pol_hel.
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR000199; Pept_3C_picorn.
DR InterPro; IPR000081; Peptidase_C3.
DR InterPro; IPR003138; Pico_P1A.
DR InterPro; IPR002527; Pico_P2B.
DR InterPro; IPR001676; RNV.
DR InterPro; IPR000605; RNA_helicase.
DR InterPro; IPR007095; RNA_pol_DS_PS.
DR InterPro; IPR001205; RNA_pol_P3D.
DR InterPro; IPR007094; RNA_pol_PSVir.
DR InterPro; IPR008975; Viral_cap_coat.
DR Pfam; PF00548; Cys-protease-3C; 1.
DR Pfam; PF02226; Pico_P1A; 1.
DR Pfam; PF00947; Pico_P2A; 1.
DR Pfam; PF01552; Pico_P2B; 1.
DR Pfam; PF00073; rhv_3.
DR Pfam; PF00680; RNA_dep_RNA_pol; 1.
DR Pfam; PF00910; RNA_helicase; 1.
PRINTS: PR00918; CALICVIRUSNS.
ProDom; PD001135; Cys_protease_3C; 1.
ProDom; PD001306; Pico_P2A; 1.
ProDom; PD001274; Pico_P2B; 1.
KW Polyprotein; Coat protein; Core protein; Transferase;
KW RNA-directed RNA polymerase; Hydrolase; Thiol protease; Myristate;
KW Lipoprotein.
FT CHAIN 2 69 COAT PROTEIN VP4.
FT CHAIN 70 340 COAT PROTEIN VP2.
FT CHAIN 341 577 COAT PROTEIN VP3.
FT CHAIN 578 888 COAT PROTEIN VP1.
FT CHAIN 889 1035 CORE PROTEIN P2A.
FT CHAIN 1036 1132 CORE PROTEIN P2B.
FT CHAIN 1133 1461 CORE PROTEIN P2C.
FT CHAIN 1462 1548 CORE PROTEIN P3A.
FT CHAIN 1549 1570 GENOME-LINKED PROTEIN VP3.
FT CHAIN 1571 1753 PICORNAIN 3C.
FT CHAIN 1754 2214 RNA-DIRECTED RNA POLYMERASE.
FT LIPID 2 N-myristoyl glycine (in host) (By
FT ACT_SITE 1717 1717 PROTEASE (POTENTIAL).
FT ACT_SITE 1731 1731 PROTEASE (POTENTIAL).
SQ SEQUENCE 2214 AA; 247212 MW; E0DD74569E1B22B8 CRC64;
Alignment Scores:
Pred. No.: 1.51 Length: 2214
Score: 95.00 Matches: 70
Percent Similarity: 35.05% Conservative: 46
Best Local Similarity: 21.15% Mismatches: 110
Query Match: 4.64% Indels: 105
DB: 1 Gaps: 17
US-10-030-529A-1 (1-1168) x POLG_CXA24 (1-2214)
QY 127 AGTGGGATTAGCTTGTTCTACTATTACACAATGGCTCAGCAGCGCAAG----- 178
Db 761 SerAsnProSerValPheTyrThrTyr-----GlySerAla-ProProArgIleSerIl 778
QY 179 -----TTTGTCTGGAGTAGTCTTCTTTGTATAGCTAGTATGAC-----TATGCTAAG----- 226
Db 778 eProTyrValGlyIleAlaAsnAlaTyrSerHisPheTyrAspGlyPheAlaArgValPr 798
QY 227 -----GGTAAATGGACTTGGTCTAA 246
Db 798 oLeuLysAspGluThrValAspSerGlyAspThrTyrTyrGlyLeuValThrIleAsnAs 818
QY 247 TGAAGCGCGTTTCGATATTAAAGTCCAGCGGATTAAATGAAGCCAAAGAAATGGATTTC 306
Db 818 pPheGlyThrLeuAlaValArgValValAsn---GluPheAsnProAlaArgIleIleSe 837
QY 307 TAAACAGGCTACTTATCTTGAATTACAGCAT-----TATATGCTTATATCTCTGT 357
Db 837 xLysIleArgValTyrMetLysProLysHisValArgCysTrpCysProArgProAr 857
QY 358 TCTCGTGACATAT---GCTCTCGCGTTTCTCTAGCCCTACTACTGTATATCCGATGTC 414
Db 857 gAlaValProTyrArgGlyGluGlyValAspPheLysGlnAspSerIleThrProLeuIl 877
QY 415 TGATCTCTGATCAACTT-----GGAATAAATCGGCAGCAGCTG----- 451
Db 877 eAlaValGluAsnIleAsnThrPheGlyGlyPheGlyHisGlnAsnMetAlaValTyrVa 897
QY 452 -----AAATGGAATTGTATAGTTATTTTAACGATTTTAAACACACACATTTTAAATT 501
Db 897 lAlaGlyTyrLysIleCysAsnTyrHisLeuAlaThrProGluAspHisAspAlaVa 917
QY 502 AAAAGTTCTTGATGACGT-----ATTCCAAAAATAACAAATATTGATAC 549
Db 917 lArgValLeuTrpAsnArgAspLeuMetIleValSerSerArgAlaGlnGlySerAspTh 937
QY 550 TATAAGTAAATATTACTAGCACTGGGTACTTAT----- 583
Db 937 rIleAlaArgCysAsnCysArgThrGlyValTyrTyrCysLysSerMetLysLysTyrTy 957

```

```
QY 584 -----TTAGATGATTTCTTATCGTATGATGAACAAAATACACATATATCAA 630
      :::::::::::::::::::: |||||
Db 957 rProValThrValThrGluProThrPheGlnThrMetGluAlaAsnAspTyrTyrProAl 977
      :::::::::::::::::::: |||||
QY 631 TAAGTTGTCTAAAGAAATTCACAACTGGTTTA----- 661
      :::::::::::::::::::: |||||
Db 977 aArgTyrGlnThrHisMetLeuLeuMetGlyPheAlaGluProGlyAspCysGlyG1 997
      :::::::::::::::::::: |||||
QY 662 -----GCCAACCAATCGCATTCGTCTATGTTAGTGCAA-----CCAAATGGTGAGG 708
      :::::::::::::::::::: |||||
Db 997 yIleLeuArgCysAsnHisGlyValMetGlyIleValThrAlaGlyGlyAsnGlyIle-- 1016
      :::::::::::::::::::: |||||
QY 709 CAAACACGAGCGTTCTCGTCGGGTAGGAGTTATAGAGAT-----AAAAC 753
      :::::::::::::::::::: |||||
Db 1017 -----ValAlaPheAlaAspIleArgAspLeuTrpValTyrGluGluG1 1031
      :::::::::::::::::::: |||||
QY 754 TGCAATTACCATTCGTGTCGGCTACGCAATCTACTGATCGCTTACCGCTAAAGCGGGTGT 813
      :::::::::::::::::::: |||||
Db 1031 uAlaMetGluGlnGlyIleThrSerTyrIleGlu----- 1042
      :::::::::::::::::::: |||||
QY 814 AGCGTTCAATACCTACAAATGCGCGCATGCTTATGCTGCTCTCTGTTGTTATGAATCTTA 873
      :::::::::::::::::::: |||||
Db 1043 -----SerLeuGlyAlaAlaPheGlySerGlyPhe-- 1052
      :::::::::::::::::::: |||||
QY 874 ATCATTACGTTTAATCACTAATCGTTTGGT 904
      :::::::::::::::::::: |||||
Db 1053 -----ThrAsnGlnIleGly 1057
      :::::::::::::::::::: |||||

RESULT 11
ID_Y946_ARCFU STANDARD; PRT; 1036 AA.
AC O29316;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein AF0946 precursor.
GN AF0946.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
OC Archaeoglobaceae; Archaeoglobus.
OX NCBI TaxID=2234;
RN [1]_
RP SEQUENCE FROM N.A.
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE=98049343; PubMed=9389475;
RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA Ketchum K.A., Dodson R.J., Winn M., Hickey E.K., Peterson J.D.,
RA Richardson D.L., Kerlavage A.R., Graham D.E., Kyriades N.C.,
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
RA Cotton M.D., Spriggs T., Artach P., Kaine B.P., Sykes S.M.,
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
RT reducing archaeon Archaeoglobus fulgidus.";
RL Nature 390:364-370(1997).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE001039; AAB90312.1; -.
DR PIR; B69368; B69368.
DR TIGR; AF0946; -.
KW Hypothetical protein; Transmembrane; Signal; Complete proteome.
```

```
FT SIGNAL 1 24 POTENTIAL.
FT CHAIN 25 1036 HYPOTHETICAL PROTEIN AF0946.
FT TRANSEM 1011 1033 POTENTIAL.
FT DOMAIN 968 1006 THR-RICH.
SQ SEQUENCE 1036 AA; 112425 MW; CA9FCE0932F9899F CRC64;

Alignment Scores:
Pred. No.: 1.63 Length: 1036
Score: 94.50 Matches: 77
Percent Similarity: 30.98% Conservatives: 46
Best Local Similarity: 19.40% Mismatches: 135
Query Match: 4.61% Indels: 139
DB: 1 Gaps: 15

US-10-030-529A-1 (1-1168) x Y946_ARCFU (1-1036)
QY 23 TTTAATGTAAGTGAAGTAAGAAAATAAATCTATATTTTACAATCAAGATTGACAATATAT 82
      ||| ::| ||| ::| ||| ::| ||| ::| ||| ::| ||| ::| ||| ::| ||| ::|
Db 535 PheThrIleAspThrAsnLeuLysIleAsnSerGluValLysPheLysIleAspAspLeu 554
      ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
QY 83 -----TTA 85
      ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
Db 555 AlaPheCysSerGlyAspProAspCysAspValLysGluLysThrTyrTyrValAspAla 574
      ||| ::| ||| ::| ||| ::| ||| ::| ||| ::| ||| ::| ||| ::| ||| ::|
QY 86 CTTAATCAGGTGATTATGAAAATT-----AAA 112
      ||| ::| ||| ::| ||| ::| ||| ::| ||| ::| ||| ::| ||| ::| ||| ::|
Db 575 LeuGlyAspValValIleLysLeuAspValAsnThrGluAlaProLeuThrAspTyrLys 594
      ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
QY 113 TGTTAGTTCGCGTAGTGGGATTAGCTTGTCTACT----- 148
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 595 PheThrAlaGluIleProGlyLeuGlyIleSerThrAspIleArgValSerValValLys 614
      ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
QY 149 -----ATTACAACAATCGCTCAGCAGCGCCAAAGTTTGTGGAGTATCTTCT 196
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 615 GlnThrLeuAspIleSerValAspArgThrThrValProArgGlyGlyAspVal 634
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 197 TTGTATAGCTATGATGATGCTATGTAAGGTGTAATGAGCTTGTCTTAATGAGCGGT 256
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 635 ThrGlySerSerThrAlaAsp-----GlyValTyrIleTyrAlaSerAspSerGly 651
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 257 -----TTCGATATTAAGTCCAGGGATTAAA 283
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 652 ValPheThrValGlyAspThrProValProAspValAspLeuLysThrLysGlySerLys 671
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 284 -----ATGACGCAAAAGAA-----TGG 301
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 672 IleAsnThrThrAspValProTyrMetGluProAspAspAsnAspAsnAspPheGln 691
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 302 ATTTCTAAACAGGCTACTTATCTTGAATTACAGCATTATATGCTTACTCTCTTCTC 361
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 692 IleSerValAsnValThrGlyValGluThrGlyThrTyrTyrLeuTyr----- 707
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 362 GTGACATATGCTCTC-----GGCGTTTCTTCAGCCCTATA 397
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 708 ---PheTyrAlaProAlaAsnIleSerValValAspLysAlaSerAspProGlnLysIle 726
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 398 CTGTTATATCCGATGCTGATCTGATCACTTGAATA----- 436
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 727 IleAlaValThrValThrAspProGlnIleValGluValThrAlaProSerLysValPro 746
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 437 -----AATCGGCGCAGCAGCTGAAA 454
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 747 TyrGlnSerLysPheGluValSerValLeuThrAspProGlyAspArgAspAsnValGlu 766
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 455 TTGAATTGTAT----- 466
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 767 ValArgLeuValLeuSerGlyProAsnValArgAspThrThrValAlaAspTrpAlaSer 786
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 467 -----AGTTATTTTAAACGATTTAAGACACAGATTTTAAATTAAAGTCTTCTGAT 514
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 797 ValAspThrAsnAsnTyrPheAsn-----LysThrValAsp 798
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 515 GCACGT---ATTTCAAAAATAAAACAAATATTTGATCTATAAGTAATAATATTACTAGAA 571
```


Db 201 GlyValGlyLysThrSerValSerAlaAlaValGlyGlyTyrArgAspLysThrAlaLeu 220

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: May 13, 2004, 08:17:32 ; Search time 64 Seconds
(without alignments)
10312.984 Million cell updates/sec

Title: US-10-030-529a-1

Perfect score: 2048

Sequence: 1 ataatactgattgacatt.....aagccgttaaaaggtcgagc 1168

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 3172214

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp
-Q=/cgn2_1/USFTO_spool_p/US10030529/runat_13052004_081451_2590/app_query.fasta_1.1351
-DB=A_Geneseq_29Jan04 -Qfmt=fastan -SUFFIX=rag -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10030529 @CGN 1.1 81 @runat_13052004_081451_2590 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSFLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : A_Geneseq_29Jan04.*
1: Geneseqp1980s.*
2: Geneseqp1990s.*
3: Geneseqp2000s.*
4: Geneseqp2001s.*
5: Geneseqp2002s.*
6: Geneseqp2003as.*
7: Geneseqp2003bs.*
8: Geneseqp2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1342	65.5	257	4 AAB31713	Aab31713 Amino aci
2	1342	65.5	257	4 AAB31705	Aab31705 Amino aci
3	1342	65.5	257	4 AAB31709	Aab31709 Amino aci
4	1299.5	63.5	264	4 AAB31706	Aab31706 Amino aci
5	1299.5	63.5	264	4 AAB31707	Aab31707 Amino aci
6	1289	62.9	271	4 AAB31711	Aab31711 Amino aci
7	1283	62.6	273	4 AAB31712	Aab31712 Amino aci
8	1259	61.5	242	4 AAB31708	Aab31708 Amino aci
9	1205	58.8	263	4 AAB31710	Aab31710 Amino aci
10	218	10.6	323	6 ABU07927	Abu07927 Neisseria

11	217	10.6	355	6 ABU07925	Abu07925 Neisseria
12	217	10.6	357	6 ABU07926	Abu07926 Neisseria
13	217	10.6	362	6 ABU07915	Abu07915 Neisseria
14	217	10.6	364	3 AAY75736	Aay75736 Neisseria
15	217	10.6	364	4 AAU27562	Aau27562 Neisseria
16	217	10.6	364	6 ABU07918	Abu07918 Neisseria
17	214	10.4	325	6 ABU07928	Abu07928 Neisseria
18	209	10.2	391	6 ABU07921	Abu07921 Neisseria
19	209	10.2	393	6 ABU07922	Abu07922 Neisseria
20	209	10.2	398	6 ABU07916	Abu07916 Neisseria
21	209	10.2	400	6 ABU07919	Abu07919 Neisseria
22	204	10.0	405	6 ABU07917	Abu07917 Neisseria
23	204	10.0	405	6 ABU07923	Abu07923 Neisseria
24	204	10.0	407	6 ABU07924	Abu07924 Neisseria
25	204	10.0	407	6 ABU07920	Abu07920 Neisseria
26	204	10.0	645	4 AAE10036	Aae10036 N. mening
27	204	10.0	645	4 AAU27601	Aau27601 Neisseria
28	204	10.0	648	4 AAE10028	Aae10028 N. mening
29	204	10.0	648	4 AAU27579	Aau27579 Neisseria
30	204	10.0	793	4 AAE10016	Aae10016 N. mening
31	204	10.0	793	4 AAU27571	Aau27571 Neisseria
32	204	10.0	806	4 AAE10033	Aae10033 N. mening
33	204	10.0	806	4 AAE10035	Aae10035 N. mening
34	204	10.0	806	4 AAU27598	Aau27598 Neisseria
35	204	10.0	806	4 AAU27600	Aau27600 Neisseria
36	204	10.0	858	4 AAE10019	Aae10019 N. mening
37	204	10.0	858	4 AAU27574	Aau27574 Neisseria
38	204	10.0	1444	4 AAE10037	Aae10037 N. mening
39	204	10.0	1444	4 AAU27602	Aau27602 Neisseria
40	204	10.0	1447	4 AAE10024	Aae10024 N. mening
41	204	10.0	1447	4 AAU27577	Aau27577 Neisseria
42	193	9.4	610	2 AAU68206	Aaw68206 M. catarr
43	193	9.4	624	2 AAU68204	Aaw68204 M. catarr
44	193	9.4	889	2 AAU68208	Aaw68208 M. catarr
45	190.5	9.3	573	2 AAU68202	Aaw68202 M. catarr

ALIGNMENTS

RESULT 1
AAB31713
ID AAB31713 standard; protein; 257 AA.
XX
AC AAB31713;
XX
DT 30-APR-2001 (first entry)
XX
DE Amino acid sequence of the DsrA protein from strain OF406.
XX
KW DsrA protein; outer membrane protein; serum resistance; vaccine;
XX chancroid disease; genital ulcer disease.
XX
OS Haemophilus ducreyi.
XX
PN WO200104138-A1.
XX
PD 18-JAN-2001.
XX
PF 07-JUL-2000; 2000WO-US018834.
XX
PR 09-JUL-1999; 99US-0143257P.
XX
PA (UYN-C-) UNIV NORTH CAROLINA.
PA (ELKI/) ELKINS C.
XX
PI Elkins C;
XX
DR WPI: 2001-138311/14.
DR N-PSDB; AAP25270.
XX
PT Novel purified DsrA protein of Haemophilus ducreyi which confers serum
PT resistance to the bacteria used to produce vaccines that induce immune
PT response against the bacteria in subject at risk of developing chancroid.

Db 1 MetLysIleLysCysLeuValAlaValGlyLeuAlaCysSerThrIleThrMet 20
 QY 161 GCTCAGACGCGCCAAAGTTGCTGGAGTATCTCTTTGTATAGCTATGAGTACTAT 220
 Db 21 AlaGlnGlnProProlLysPheAlaGlyValSerSerLeuTyrSerTyrGluTyrAspTyr 40
 QY 221 GGTAAAGGTTAAATGCACTTGGTCTTAATGAAGCGGTTTCGATATATAAGTCCAGGATT 280
 Db 41 GlyLysGlyLysThrThrTrpSerAsnGluGlyPheAspIleLysValProGlyIle 60
 QY 281 AAATGAAGCCAAAGAAATGATTTCTTAACAGGCTTCTATCTTGAATACACATAT 340
 Db 61 LysMetLysProlLysGluTrpIleSerLysGlnAlaThrTyrLeuGluLeuGlnHisTyr 80
 QY 341 ATGCCTTATATCTCTGCTCGTACATATGCTCTCGGCTTCTCTAGCCCTATAGT 400
 Db 81 MetProTyrThrProValLeuValThrTyrAlaProGlyValSerProSerProIleLeu 100
 QY 401 TTATATCCGATGCTGATCTGATCAACTTGAATAAATCGGACGAGCTGAATTCGAAT 460
 Db 101 LeuTyrProMetSerAspProAspGlnLeuGlyIleAsnArgGlnGlnLeuLysLeuAsn 120
 QY 461 TTGTATAGTTATTTAAACGATTTAAAGACACGATTTTAAATTTAAAGTTCTTGATGCACGT 520
 Db 121 LeuTyrSerTyrPheAsnAspLeuArgHisAspPheLysLeuLysValLeuAspAlaArg 140
 QY 521 ATTTCCAAAATAAACAATATTTGATCTATAGTAAATTTTACTAGAACTGGGTACT 580
 Db 141 IleSerLysAsnLysGlnAsnIleAspThrIleSerLysTyrLeuLeuGluLeuGlyThr 160
 QY 581 TATTAGATGATCTTATCGTATCATGGAACAAATACACATAATATCAATAGTCTGCT 640
 Db 161 TyrLeuAspAspSerTyrArgMetGluGlnAsnThrHisAsnIleAsnLysLeuSer 180
 QY 641 AAAGAATTGCAACTGGTTAGCCAACTATGATTCAGCATTTCTATGTTAGTGCACCAAT 700
 Db 181 LysGluLeuGlnThrGlyLeuAlaAsnGlnSerAlaLeuSerMetLeuValGlnProAsn 200
 QY 701 GGTGTAGCAAAACGACGCTTCTGCTGCGTAGGAGGTTATAGAGATAAATCGCATTA 760
 Db 201 GlyValGlyLysThrSerValSerAlaAlaValGlyGlyTyrArgAspLysThrAlaLeu 220
 QY 761 GCATTCGCTGCGCTCACGATTCATGATCGCTTTTACCGCTTAAGCGGTTAGCGCTC 820
 Db 221 AlaIleGlyValGlySerArgIleThrAspArgPheThrAlaLysAlaGlyValAlaPhe 240
 QY 821 AATACCTACAAATGCGGCATGCTTATGCTGCTTCTGTTGTTATGAATTC 871
 Db 241 AsnThrTyrAsnGlyGlyMetSerTyrGlyAlaSerValGlyTyrGluPhe 257

RESULT 3
 AAB31709
 ID AAB31709 standard; protein; 257 AA.
 AC AAB31709;
 XX
 XX 30-APR-2001 (first entry)
 DT
 XX
 XX Amino acid sequence of the DsrA protein from strain CIP542 (CDC).
 DE
 XX
 XX DsrA protein; outer membrane protein; serum resistance; vaccine;
 KW chancroid disease; genital ulcer disease.
 KW
 XX Haemophilus ducreyi.
 OS
 XX
 XX W0200104138-A1.
 PN
 XX
 XX 18-JAN-2001.
 PD
 XX
 XX 07-JUL-2000; 2000WO-US018834.
 PF
 XX
 XX 09-JUL-1999; 99US-0143257P.
 PR
 XX
 XX

(UNCL-) UNIV NORTH CAROLINA.
 (ELKI/) ELKINS C.

Elkins C;

WPI; 2001-138311/14.

N-PSDB; AAF25266.

Novel purified DsrA protein of Haemophilus ducreyi which confers serum resistance to the bacteria used to produce vaccines that induce immune response against the bacteria in subject at risk of developing chancroid.
 Claim 7; Page 55; 80pp; English.

The present sequence represents a DsrA 30 kDa protein of Haemophilus ducreyi. DsrA is an outer membrane protein that confers serum resistance to the bacteria. DsrA antisense oligonucleotides are useful for detecting a polynucleotide which encodes DsrA in a biological sample. The DsrA polypeptide is used to produce vaccine compositions, which are useful for inducing a protective immune response in a subject at risk of developing Haemophilus ducreyi infection e.g. chancroid, a genital ulcer disease transmitted by sexual contact. DsrA, its catalytic or immunogenic fragments or peptides can be used for screening libraries of compounds in a variety of drug screening techniques. The proteins and peptides may be used as antigens in immunoassays for detection of Haemophilus ducreyi in various tissues and body fluids, e.g. blood, spinal fluid, sputum etc. . . The nucleic acids are useful for the preparation of DsrA proteins

SQ Sequence 257 AA;

Alignment Scores:
 Pred. No.: 7,02e-148 Length: 257
 Score: 1342.00 Matches: 257
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 65.53% Indels: 0
 DB: 4 Gaps: 0

US-10-030-529A-1 (1-1168) x AAB31709 (1-257)

QY 101 ATGAAATTAATAATGTTTAGTTGCGTAGTGGGATAGCTTGTCTACTATTACAAATG 160
 Db 1 MetLysIleLysCysLeuValAlaValGlyLeuAlaCysSerThrIleThrMet 20
 QY 161 GCTCAGACGCGCCAAAGTTGCTGGAGTATCTCTTTGTATAGCTATGAGTACTAT 220
 Db 21 AlaGlnGlnProProlLysPheAlaGlyValSerSerLeuTyrSerTyrGluTyrAspTyr 40
 QY 221 GGTAAAGGTTAAATGCACTTGGTCTTAATGAAGCGGTTTCGATATATAAGTCCAGGATT 280
 Db 41 GlyLysGlyLysThrThrTrpSerAsnGluGlyPheAspIleLysValProGlyIle 60
 QY 281 AAATGAAGCCAAAGAAATGATTTCTTAACAGGCTTCTATCTTGAATACACATAT 340
 Db 61 LysMetLysProlLysGluTrpIleSerLysGlnAlaThrTyrLeuGluLeuGlnHisTyr 80
 QY 341 ATGCCTTATATCTCTGCTCGTACATATGCTCTCGGCTTCTCTAGCCCTATAGT 400
 Db 81 MetProTyrThrProValLeuValThrTyrAlaProGlyValSerProSerProIleLeu 100
 QY 401 TTATATCCGATGCTGATCTGATCAACTTGAATAAATCGGACGAGCTGAATTCGAAT 460
 Db 101 LeuTyrProMetSerAspProAspGlnLeuGlyIleAsnArgGlnGlnLeuLysLeuAsn 120
 QY 461 TTGTATAGTTATTTAAACGATTTAAAGACACGATTTTAAATTTAAAGTTCTTGATGCACGT 520
 Db 121 LeuTyrSerTyrPheAsnAspLeuArgHisAspPheLysLeuLysValLeuAspAlaArg 140
 QY 521 ATTTCCAAAATAAACAATATTTGATCTATAGTAAATTTTACTAGAACTGGGTACT 580
 Db 141 IleSerLysAsnLysGlnAsnIleAspThrIleSerLysTyrLeuLeuGluLeuGlyThr 160
 QY 581 TATTAGATGATCTTATCGTATCATGGAACAAATACACATAATATCAATAGTCTGCT 640

|||||
161 TyrLeuAspSerTyrArgMetGluGlnAsnThrHisAsnLysLeuSer 180
Qy 641 AAAGAATTGCAACTGTTAGTCAACCAATCAGCATGCTATGTTAGTGCACCAAT 700
Db 181 LysGluLeuGlnThrGlyLeuAlaAsnGlnSerAlaLeuSerMetLeuValGlnProAsn 200
Qy 701 GGTCTAGGCAAAACGAGCGTTTCTGCTGCGGTAGGAGTTATAGAGATAAACTGCATTA 760
Db 201 GlyValGlyLysThrSerValSerAlaAlaValGlyTyrArgAspLysThrAlaLeu 220
Qy 761 GCCATTGGTGTGCGCTCACGCAATTACTGATCGCTTTACCGCTAAAGCGGGTGTAGCGTTC 820
Db 221 AlaIleGlyValGlySerArgIleThrAspArgPheThrAlaLysAlaGlyValAlaPhe 240
Qy 821 AATACCTACAAATGCGGCGATGCTTATGCTGCTCTCTGTTGGTTATGAATTC 871
Db 241 AsnThrTyrAsnGlyGlyMetSerTyrGlyAlaSerValGlyTyrGluPhe 257
RESULT 4
AAB31706
ID AAB31706 standard; protein; 264 AA.
XX AC AAB31706;
XX DT 30-APR-2001 (first entry)
XX DE Amino acid sequence of the DsrA protein from strain CIPA75.
XX KW DsrA protein; outer membrane protein; serum resistance; vaccine;
XX KW chancroid disease; genital ulcer disease.
XX OS Haemophilus ducreyi.
XX FN WO200104138-A1.
XX PD 18-JAN-2001.
XX PF 07-JUL-2000; 2000WO-US018834.
XX PR 09-JUL-1999; 99US-0143257P.
XX PA (UYNC-) UNIV NORTH CAROLINA.
XX PA (ELKI/) ELKINS C.
XX FI Elkins C;
XX DR WPI; 2001-138311/14.
XX DR N-PSDB; AAF25263.
XX PT Novel purified DsrA protein of Haemophilus ducreyi which confers serum
PT resistance to the bacteria used to produce vaccines that induce immune
PT response against the bacteria in subject at risk of developing chancroid.
XX PS Claim 7; Page 53; 80pp; English.
XX CC The present sequence represents a DsrA 10 kDa protein of Haemophilus
CC ducreyi. DsrA is an outer membrane protein that confers serum resistance
CC to the bacteria. DsrA antisense oligonucleotides are useful for detecting
CC a polynucleotide which encodes DsrA in a biological sample. The DsrA
CC polypeptide is used to produce vaccine compositions, which are useful for
CC inducing a protective immune response in a subject at risk of developing
CC Haemophilus ducreyi infection e.g. chancroid, a genital ulcer disease
CC transmitted by sexual contact. DsrA, its catalytic or immunogenic
CC fragments or peptides can be used for screening libraries of compounds in
CC a variety of drug screening techniques. The proteins and peptides may be
CC used as antigens in immunoassays for detection of Haemophilus ducreyi in
CC various tissues and body fluids, e.g. blood, spinal fluid, sputum etc. .
XX CC The nucleic acids are useful for the preparation of DsrA proteins
SQ Sequence 264 AA;
Alignment Scores:

Pred. No.: 6.83e-143 Length: 264
Score: 1299.50 Matches: 253
Percent Similarity: 95.83% Conservative: 0
Best Local Similarity: 95.83% Mismatches: 4
Query Match: 63.45% Indels: 7
DB: 4 Gaps: 1
US-10-030-529A-1 (1-1168) x AAB31706 (1-264)
Qy 101 ATGAAATTAATAGTTTAGTTGCGGTAGTGGATTAGCTTGTCTACTATTACACATG 160
Db 1 MetLysIleLysCysLeuValAlaValGlyLeuAlaCysSerThrIleThrMet 20
Qy 161 GCTCAGCAGCGCCAAAGTTTCTGCTGGAGTATCTTCTTTGTATAGCTATGAGTATGACTAT 220
Db 21 AlaGlnGlnProProLysPheAlaGlyValSerSerLeuTyrSerTyrGluTyrAspTyr 40
Qy 221 GGTAAGGTAATAGGACTTCGTTCTTAATGAAGCGGGTTTCGATATATAAGTCCAGGAGTT 280
Db 41 GlyLysGlyLysTrpThrTrpSerAsnGluGlyGlyPheAspIleLysValProGlyIle 60
Qy 281 AAAATGAAGCCAAAGATCGATTCTTAAACAGCTACTTATCTTGAATTACAGCATAT 340
Db 61 LysMetLysProLysGluTrpIleSerLysGlnAlaThrTyrLeuGluLeuGlnHisTyr 80
Qy 341 ATGCTTATATCTCTGTTCTCGTGACATATATGCTCTGCGGTTTCTCTAGCCCTATATCTG 400
Db 81 MetProTyrThrProValLeuValThrTyrAlaHisAspValProProSerSerIleLeu 100
Qy 401 TTATATCCGATGTCTGATCCTGATCAACTTGGATAAATCGGCAGCAGCTGAATTAAT 460
Db 101 LeuTyrProMetSerAspProAspGlnLeuGlyIleAsnArgGlnGlnLeuLysLeuAsn 120
Qy 461 TTGTATAGTTATTTTAACGATTTTAAGACACAGATTTTAAATTAAGTCTTGTATGCACGT 520
Db 121 LeuTyrSerTyrPheAsnAspLeuArgHisAspPheLysLeuLysValLeuAspAlaArg 140
Qy 521 ATTTCCAAAAATAACAAAAATATTGATACATATAAGTAAATATTACTAGAACTGGGTACT 580
Db 141 IleSerLysAsnLysGlnAsnIleAspThrIleSerLysTyrLeuLeuGluLeuGlyThr 160
Qy 581 TATTAGATGATTTTATCGTATGATGAAACAA-----AATACA 619
Db 161 TyrLeuAspAspSerTyrArgMetMetGluGlnAsnThrHisAsnIleAsnLysAsnThr 180
Qy 620 CATATATCAATAAGTTGTTCTTAAGAATTGCAAACTGGTTTAGCAACCAATCAGCATG 679
Db 181 HisAsnIleAsnLysLeuSerLysGluLeuGlnThrGlyLeuAlaAsnGlnSerAlaLeu 200
Qy 680 TCTATGTTAGTGAACCAAAATGGTGTAGGCAAAACGAGCGTTTCTGCTCGGTAGGAGGT 739
Db 201 SerMetLeuValGlnProAsnGlyValGlyLysThrSerValSerAlaAlaValGlyGly 220
Qy 740 TATAGATATAAACTGCATTAGCCATTGGTGTGCGCTCACGCATTACTGATCGCTTTACC 799
Db 221 TyrArgAspLysThrAlaLeuAlaIleGlyValGlySerArgIleThrAspArgPheThr 240
Qy 800 GCTAAACGGGTGTAGGTTCAATACCTACATGCGCGCATGCTTATGGTCTCTGTT 859
Db 241 AlaLysAlaGlyValAlaPheAsnThrTyrAsnGlyGlyMetSerTyrGlyAlaSerVal 260
Qy 860 GGTTCATCAATTC 871
Db 261 GlyTyrGluPhe 264
RESULT 5
AAB31707
ID AAB31707 standard; protein; 264 AA.
XX AC AAB31707;
XX DT 30-APR-2001 (first entry)
XX

```
Db 799 LeuArgaspLeuAlaLysAsnLysLeuAsnLeuAspAlaLeuGluProGlyLeuTyrVal 818
Qy 572 CTGGTACTATTAGATGATCTTATCGTATGATGGAACAAATACACATATATCAAT 631
Db 819 LeuThrAlaGluLeuArgPheLysGlnSerLeuGlyGlyGlyValAspSerGluAsp 838
Qy 632 AAGTTGTCTAAAGAAATTCACAACTGGTTAGCCCAACCAATCAGCATTTCTATGTTAGTG 691
Db 839 LysLeuLeuLeuLeuGlyLeuThrPheGluValAspValAsnThrProValValle 858
Qy 692 CAACCAATGGTGTAGGACAAACAGAGCGTTCTCTCGGTGAGGAGTTATAGAGATAAA 751
Db 859 GlyAspGluLeuValValAsnLeuThrAspArgLeuGluAlaGlyTyrAspLysile 878
Qy 752 ACTGCATTAGCATTTGGTGGCTCAGCATTTACTGATCGCTTTACC----- 799
Db 879 PheValThrLeuValGlyThrAsnTyrLysValThrGlnValAlaThrLeuAsnSerGlu 898
Qy 800 GCTAAAGCGGTGTAGCGTTCAATACCTACAAATGGCGCATGCTTATGGT 850
Db 899 GlyLysAlaThrValThrPheGluThrTyr-----GlyMetSerAlaGly 913
```

RESULT 12

```
CMC2_CABEL
ID_CMC2_CABEL STANDARD; PRT; 588 AA.
AC Q20759;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE 10-OCT-2003 (Rel. 42, Last annotation update)
DE Probable calcium-binding mitochondrial carrier F55A11.4.
GN F55A11.4
OS Caenorhabditis elegans.
OC Rukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Kershaw J.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: Calcium-dependent mitochondrial solute carrier (By
CC similarity).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Mitochondrial
CC inner membrane (By similarity).
CC -1- SIMILARITY: Belongs to the mitochondrial carrier family.
CC -1- SIMILARITY: Contains 4 EF-hand calcium-binding domains.
CC -1- SIMILARITY: Contains 3 Solcar repeats.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z72511; CAA96658.3; -
DR PIR; T22688; T22688.
DR WormPep; F55A11.4; CE05946.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR002067; Mit_carrier.
DR InterPro; IPR001993; Mitoch_carrier.
DR Pfam; PF00036; efhand; 3.
DR Pfam; PF00153; mito_carr; 3.
DR PRINTS; PR00926; MITOCARRIER.
DR ProDom; PD000012; EF-hand; 1.
DR SMART; SM00054; EFH; 3.
DR PROSITE; PS00018; EF_HAND; 2.
DR PROSITE; PS00920; SOLCAR; 3.
KW Hypothetical protein; Mitochondrion; Inner membrane; Repeat;
KW Transmembrane; Transport; Calcium-binding.
FT TRANSMEM 252 269 1 (POTENTIAL).
```

```
FT TRANSMEM 307 326 2 (POTENTIAL).
FT TRANSMEM 352 365 3 (POTENTIAL).
FT TRANSMEM 403 422 4 (POTENTIAL).
FT TRANSMEM 446 463 5 (POTENTIAL).
FT TRANSMEM 504 523 6 (POTENTIAL).
FT CA_BIND 52 63 EF-HAND 1.
FT CA_BIND 86 97 EF-HAND 2.
FT DOMAIN 123 133 ANCESTRAL CALCIUM SITE 3.
FT CA_BIND 153 164 EF-HAND 4.
FT REPEAT 246 332 SOLCAR 1.
FT REPEAT 342 428 SOLCAR 2.
FT REPEAT 440 529 SOLCAR 3.
SQ SEQUENCE 588 AA; 66336 MW; 68DDF60923D8697D CRC64;
```

Alignment Scores:

```
Pred. No.: 2.4 Length: 588
Score: 92.50 Matches: 48
Percent Similarity: 41.59% Conservative: 41
Best Local Similarity: 22.43% Mismatches: 72
Query Match: 4.52% Indels: 53
DB: 1 Gaps: 12
```

US-10-030-529A-1 (1-1168) x CMC2_CABEL (1-588)

```
Qy 245 AATGAAGCGGTTTCGATATTTAAAGTCCAGGAGTTAAATGAAGCCAAAAGAATGGATT 304
Db 89 AsnAspGlyThrIleAspIleArg----- 96
Qy 305 TCTAAACAGCGCTACTTATCTTGAATTAACAGCATATATGCTTATATCTCTGCTCGTG 364
Db 97 -----AspLeuThrLeuAlaLeuLysHisGluThrProHisIlePro----- 110
Qy 365 ACATATGCTCTCGGCTTCTCTAGCCCTATATGCTTATATCGCATTCGATCCGAT 424
Db 111 -----AlaAsnLeuAlaProValIleMetSerLysMetSer---ProAsp 124
Qy 425 CAACTTGAATAAATCGGCAGCAGCTGAAATTTGAATTTGATAGTTATTTTAAAC---GAT 481
Db 125 AspGluGly-----ArgValAspPheTyrSerPheSerTyrVal 138
Qy 482 TTAAGCACAGATTTTAAATTAAGTTCTTGTGACGAGTATTTCCAAAATAAACAATAAT 541
Db 139 LeuGluAsnGluGlnLysLeuAlaGluMetPheAlaAspMetAspArgAsnHisAspGly 158
Qy 542 -----ATTGATACTATA-----AGTAATATTTACTAGAACCTGGTACTTATTTAGATGAT 592
Db 159 LeuValAspValValGluMetLysAsnTyrCysLysAspIleGlyValProLeuAspAsp 178
Qy 593 TCTATCGTATGATGGAACAAATACACATAATCAATCAATGATGTTCTTAAAGAATTGCAA 652
Db 179 -----HisLysAlaGlnHisIleValAlaAsnLysMetAsp-----Gln 190
Qy 653 ACTGTTTTCAGCAACCAATCAGCATTTGCTCT-----ATGTTAGTGCACCAACCAAT 700
Db 191 ThrGlySerAlaValAspValAspLeuLysGluPheGlnGluPheMetLeuTyrProSer 210
Qy 701 GGTGTAGGCAAAACGAGCGGTTTCTGTCGCGTAGGAGTTATAGAGATAAATCGCATTA 760
Db 211 -----SerAspLeuLysAspIleValAspPheTyrArgHisAsnLeuIle 226
Qy 761 GCCATTGGTGTGCGCTCAGCATTTACTGATCGCTTTACC-----GCTAAAGCGGT 811
Db 227 AspIleGlyGluAspSerGlnIleProGluAspPheSerGlnGlnGluMetGlnGluGly 246
Qy 812 GTAGCGTTCAATACCTACAAATGGCGCATGCTCTTATGGTGT 853
Db 247 IleTyrTyrPargHisLeuValAlaGlyAlaGlyAla 260
```

RESULT 13

```
MFPI_ARATH
ID_MFPI_ARATH STANDARD; PRT; 727 AA.
AC Q9LW85;
DT 16-OCT-2001 (Rel. 40, Created)
```


KW DNA replication; ATP-binding; Helicase.
 FT NP_BIND 90 97 ATP (BY SIMILARITY)
 SQ SEQUENCE 824 AA; 93286 MW; 5A5057544E06184C CRC64;

Alignment Scores:

Pred. No.: 2.43 Length: 824
 Score: 92.50 Matches: 53
 Percent Similarity: 38.75% Conservative: 40
 Best Local Similarity: 22.08% Mismatches: 94
 Query Match: 4.52% Indels: 53
 DB: 1 Gaps: 12

US-10-030-529A-1 (1-1168) x HELI_HSV6U (1-824)

QY 329 TTACAGATTAT-----ATGCCCTATCTCTGTCCTGACATATGCTCCTGGC 379
 Db 315 LeuylsHisileGluPheGlyLeuProleuylsProGluLeuileGluTyrValAspGln 334
 QY 380 ---GTTTCTCTAGCCCTATCTATATCCGATGCTGATCTGATCTGAT-----CAACTT 430
 Db 335 PheValylsProAlaSerTyrIleArgAsnPrometAsnGluileGluThrArgLeu 354
 QY 431 GGAATAATCGGACGACGCTGAATTTGATATGTTATTTTAAAGATTAAAGACAC 490
 Db 355 PheLeuSerHisAsnGluValyls-----AsnTyrPheArgSerLeuHisGlu 370
 QY 491 GATTTTAAATTAAGTCTTGATGACGATTTTCCAAATAAACAATAATTT----- 544
 Db 371 GlnValGluVal-----ThrAsnArgAsnAsnLeuPheVal 382
 QY 545 -----GATACTATAAGTAATATTTACTAGAACTGGGT 577
 Db 383 PheProValTyrPheLeuileLysAsnLysThrPheGluAspTyrLysSerGluileGly 402
 QY 578 ACTTATTAGATGTTCTTATCGTATGATGACCAAAATACACATAATATCAATAAGTTG 637
 Db 403 AsnPheSerLeuGluileGluProTyrPheLysSerAsnileHisArgLeuAsnThrTyr 422
 QY 638 TCTAAAGATTGCAAACTGGTTAGCCACCAATCAGCATGCTGTCTATGTTAGTCCACCA 697
 Db 423 SerGlnPheAlaAspGlnAspLeuSerLysThrValGlnLeuGluileValLeuGlu 442
 QY 698 AATGGT---GTAGCAAAACGAGGTTTCTGCTCGGTAGGAGGTATAGAGATAAAACT 754
 Db 443 AspGlySerValGluGluThrLeuileThrCysHisLeuLysHisIleArgAsnSer--- 461
 QY 755 GCATTAGCATGTTGTCGGCTCAGCATTAATGCTGCTTTTACCGCTTAAAGCGGTTGTA 814
 Db 462 -----SerileGlyValThrSerLysile-----LysAlaSer--- 472
 QY 815 GCGTTCAATACCTACAATGCGCGCATGCTTATGCTGCTTCTGTTGTTATGAA---TTC 871
 Db 473 -----ThrValGlyPheSerGlyThrTyrGluLysPhe 483
 QY 872 TAATCATTA-----CGTTTAACTACATCGTTTGGTTTATATAAAAGCGCTAAA 922
 Db 484 ValGluLeuLeuGlnSerAspLeuPheileGluLysThrSerCysAspGlnThrIleHis 503
 QY 923 TGTTTCTCTCCATCATTAGCCCTTTCTTATTTATCTTTGTTATAGCTTTTGGTGTATAAA 982
 Db 504 AlaTyrSerPheLeuSerGlyLeuMetPheGlyGlyMetTyrSerPheCysSerLys 523
 RESULT 15
 CHLB_CHLPT
 ID CHLB_CHLPT STANDARD; PRT; 444 AA.
 AC P37824;
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 01-OCT-1994 (Rel. 30, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Light-independent protochlorophyllide reductase subunit B
 DE (EC 1.18.-.-) (LI-POR subunit B) (DPOR subunit B) (Fragment).
 GN CHLB.
 OS Chlamydomonas pilschmannii.

OG Chloroplast.
 OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
 OC Chlamydomonadaceae; Chlamydomonas.
 OX NCBI_TaxID=3061;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94033309; PubMed=8219066;
 RA Liu X.-Q., Xu H., Huang C.;
 RT "Chloroplast chlB gene is required for light-independent chlorophyll
 accumulation in Chlamydomonas reinhardtii.";
 RL Plant Mol. Biol. 23:297-308(1993).
 CC -!- FUNCTION: Uses Mg-ATP and reduced ferredoxin to reduce ring D of
 CC protochlorophyllide (Pchl) to form chlorophyllide a (Chlide)
 CC (by similarity). This reaction is light-independent.
 CC -!- PATHWAY: Light-independent chlorophyll biosynthesis.
 CC -!- SUBUNIT: Protochlorophyllide reductase is thought to be composed
 CC of three subunits; chlL, chlN and chlB. Could form a
 CC heterotrimer of two chlB and two chlN subunits (By similarity).
 CC -!- SIMILARITY: Belongs to the chlB / bchB / bchZ family.
 DR HAMAP: MF_00353; atypical; 1
 DR InterPro: IPR000510; Oxidoreductase.
 DR Pfam: PF00148; oxidoreductase; 1.
 KW Oxidoreductase; Photosynthesis; Chlorophyll biosynthesis; Chloroplast.
 FT NON_TER 444 444
 SQ SEQUENCE 444 AA; 50058 MW; 6906E52CAAE4857F CRC64;
 Alignment Scores:
 Pred. No.: 2.64 Length: 444
 Score: 92.00 Matches: 60
 Percent Similarity: 39.23% Conservative: 42
 Best Local Similarity: 23.08% Mismatches: 98
 Query Match: 4.49% Indels: 61
 DB: 1 Gaps: 13
 US-10-030-529A-1 (1-1168) x CHLB_CHLPT (1-444)
 QY 38 AATAAGAAAGTAATCTATATTTACAATCAAGATTGACAAATTTTACTTAATGAGGTG 97
 Db 124 AsplysalaThrSerAspValileLeuAlaAspValAsnHisTyrArgValAsnGluLeu 143
 QY 98 ATTATGAAAATTAATGTTTAGTTGCGGTAGTG-----GGA 133
 Db 144 GlnAlaAlaAspArgThrLeuGluileValArgPheTyrLeuGluLysGluLysLys 163
 QY 134 TTAGCTTGTTCTACTATTACAAATGCTCAGCAGCGC---CCAAAGTTTGTGAGTA 190
 Db 164 LeuAsnThrAsnThrileProThrLysThrLysLysProSerAlaAsnileileGlyle 183
 QY 191 TCTTCTTTG---TATAGCTATGATGATGATGTAAGGTAATGACTTGTCTTAAT 247
 Db 184 PheThrLeuGlyPheHisAsnGlnHisAspCysArgGluLeuLysArgLeuLeuAsnAsn 203
 QY 248 -----GAAGCGGTTTTCGATATATAAGTCCAGGG 277
 Db 204 LeuGlyileGluValAsnGluileleProGluGlyGly-----SerValThrAsn 220
 QY 278 ATTAAGTAAGCAAGAAAGTAATGATTTCT-----AAA 310
 Db 221 LeuLysAsnLeuProHisAlaTyrPheAsnLeuValProTyrArgGluileGlyLeuMet 240
 QY 311 CAGCTACTTATCTTGAATTTACAGCATATATGCTTATCTCTGTTCTGTCGACATAT 370
 Db 241 ArgAlaValTyrLeuGluLysGluPheAsnMetProTyr-----Valaile 256
 QY 371 GCTCCT---GGCGTTTCTCCTAGCCCTATCTGTTATATCCGATGCTGATCTCGATCAA 427
 Db 257 SerProLeuGlyileleAspThrAlaValCysileArgGlu----- 270
 QY 428 CTTGGAATAAATCGCAGCAGCTGAAATGCAATTTGATAGTTATTTTAAAGATTAAAGA 487
 Db 271 -----ileGluLysileleLeuAsnAsnLeuSerLeuAsnGlyTyrGlnSerSerLeuPro 288
 QY 488 CACATTTTAAATA-----AAGTTCTTGATGACGATTTTCCAAATAAACAATAAT 541

```

Db      |||||   ::: |||::: |||||   |||||   :::
289  GluGlyHisLysLeuThrAsnGluGluLeuAsnGlyLeuSerGlnLysAsnAsnGlu--- 307
QY      |||||   ::: |||::: |||||   |||||   :::
542  ATTGATACTATAAGTAAATATTACTAGAACTGGTACTTATTAGATGATTCTTATCGT 601
Db      |||||   |||  |||  |||||   |||||   :::
308  -----ProSerLeuProGluGluHisArgGlnLeuAspGlyAlaLeuAsp 322
QY      |||||   |||  |||  |||||   |||||   |||:::
602  ATGATGGAAACAAATACATAATATC---AATAAGTTGTCTAAAGAATT----- 648
Db      |||||   |||  |||  |||||   |||||   |||:::
323  SerLysSerGlnLysThrTyrAsnPheGluAsnLysTyrIleLysGlnGlnThrArgPhe 342
QY      |||||   |||  |||  |||||   |||||   |||:::
649  -----GCAAACTGTTTAGCCAAACCAATGTCATGTTAGTGCAACCAAA 699
Db      |||||   |||  |||  |||||   |||||   |||:::
343  IleSerGlnAlaAlaTyrPheSerArgSerIleAspCysLeuAsnLeuThrAlaLysLys 362

```

Search completed: May 13, 2004, 08:26:09
 Job time : 28.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: May 13, 2004, 08:21:52 ; Search time 26.5 seconds

(without alignments)

8479.374 Million cell updates/sec

Title: US-10-030-529A-1

Perfect score: 2048

Sequence: 1 ataataacgcatcgacatt.....aagccgtaaaaggtgcggac 1168

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 566732

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO.spool.p/US10030529/runat 13052004 081453 2632/app_query.fasta_1.1351
-DB=PIR_78 -QFMT=fastan -SUFFIX=ipr -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THRT=SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptp -NORM=ext -HEA=SIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USBR=US10030529 @CGN_1_1_38 @runat 13052004 081453 2632 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSFBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

PIR_78:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	217	10.6	364	2 A81019	adhesin/invasin, p
2	186.5	9.1	455	2 S04912	yopA protein - Yer
3	186	9.1	422	2 S04911	yopA protein - Yer
4	182	8.9	434	2 S04534	invasin precursor
5	144	7.2	403	1 H64142	hypothetical prote
6	132	6.4	155	2 AH3277	immunoglobulin-bin
7	130	6.5	227	1 H65039	hypothetical prote
8	129	6.5	227	2 G85907	probable transport
9	129	6.5	398	2 C91063	hypothetical prote
10	119.5	5.8	413	2 AC0834	probable membrane
11	119.5	5.5	658	2 AE0110	probable surface p
12	112.5	5.5	364	2 AE0169	probable exported
13	109.5	5.3	338	2 D90697	adhesin/invasin-li
14	109.5	5.3	338	2 G85547	adhesin/invasin-li

C 15	107	5.4	427	2 AF0400	probable membrane
C 16	104	5.2	426	2 G82309	probable hemolysin
C 17	102.5	5.0	498	2 H69623	flagellar hook-ass
C 18	102.5	5.0	622	2 AF0169	probable exported
C 19	99.5	4.9	466	2 H98045	hypothetical prote
C 20	99.5	4.9	831	2 S76235	hypothetical prote
C 21	99.5	4.9	1695	2 A56921	kinesin family pro
C 22	99	4.8	795	1 D64226	endopeptidase La (
C 23	98.5	4.8	466	2 A95179	hypothetical prote
C 24	97.5	4.8	729	2 T45780	sugar transporter-
C 25	95	4.6	276	2 S05856	gene Atrase6 intro
C 26	95	4.6	2214	1 A48548	genome polyprotein
C 27	94.5	4.6	301	2 S44255	hypothetical prote
C 28	94.5	4.6	1036	2 B69368	hypothetical prote
C 29	93.5	4.6	163	2 T28380	ORF MSV219 hypothe
C 30	92.5	4.5	588	2 T22888	hypothetical prote
C 31	92.5	4.5	940	2 S19702	fibronectin-bindin
C 32	92	4.5	316	2 G81045	pilin gene inverti
C 33	92	4.5	335	2 AF1125	phosphate transpor
C 34	92	4.5	1338	2 T02206	hypothetical prote
C 35	91	4.4	659	2 AB3031	conserved hypotet
C 36	91	4.4	765	2 E98254	hypothetical prote
C 37	91	4.4	2715	2 T13049	eyelid - fruit fly
C 38	90.5	4.4	340	2 H81348	probable periplasm
C 39	90.5	4.4	819	1 B72128	endopeptidase La (
C 40	90.5	4.4	819	2 E86494	Lon ATP-dependent
C 41	90.5	4.4	827	2 AD1311	penicillin-binding
C 42	90.5	4.4	1212	2 A96971	cobalamine-depende
C 43	90.5	4.4	1226	1 S65593	adenosine deaminas
C 44	90	4.4	870	2 S65158	PAL1 protein - yea
C 45	90	4.4	1328	1 S04273	retrovirus-related

ALIGNMENTS

RESULT 1

A81019
adhesin/invasin, probable NMB1994 [imported] - Neisseria meningitidis (strain MC58 serog
C;Species: Neisseria meningitidis
C;Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001
C;Accession: A81019
R;Tettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A.
Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.;
xi, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Maignani, V.; Pizza, M.
Science 287, 1809-1815, 2000
A;Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Ver
A;Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.
A;Reference number: A81000; MUID:20175755; PMID:10710307
A;Accession: A81019
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-364 <TET>
A;Cross-references: GB:AE002548; GB:AE002098; NID:g7227249; PIDN:AAF42321.1; PID:g722725
A;Experimental source: serogroup B, strain MC58
C;Genetics:
A;Gene: NMB1994

Alignment Scores:			
Pred. No.:	1.47e-10	Length:	364
Score:	217.00	Matches:	52
Percent Similarity:	53.79%	Conservative:	19
Best Local Similarity:	39.39%	Mismatches:	37
Query Match:	10.60%	Indels:	24
DB:	2	Gaps:	3

US-10-030-529A-1 (1-1168) x A81019 (1-364)

Qy	479	GATTTAAGACACGATTTAAATGAAGTCTTGATGCAGCATTTCCAAAATAACAA	538
Db	256	AspilleysAlaAspAlaThrAsnLysAlaAsp-----IleAlatysAsnSerAla	273
Qy	539	AAATATTGATCTAAAGTAAATATTACTAGAACTGGTACTTATTATTAGATGATCTTAT	598

```
Db 274 ArgileAspSerLeuAspLys----- 280
Qy 599 CGTATGATGGACAAATACACATAATATCAATAGTTGTCTAAAGATTTCGAACATGGT 658
Db 281 -----AsnValAlaAsnLeuArgLysGluThrArgGlnGly 292
Qy 659 TTAGCCAAATCAGCATTTCTATGTAGTCAACCAATAGTGTAGGCAAAACGAGC 718
Db 293 LeuAlaGluGlnAlaAlaLeuSerGlyLeuPheGlnProTyrAsnValGlyArgPheAsn 312
Qy 719 GTTCTGTCGGTAGGAGGTATAGACATAAACTGCATTAGCATTGGTGGCTCA 778
Db 313 ValThrAlaAlaValGlyGlyTyrLysSerGluSerAlaValAlaIleGlyThrGlyPhe 332
Qy 779 CGCATTACTGATCGCTTTACCGCTAAACGGGTGTAGCGTTCAATACCTACCAATGGC--- 835
Db 333 ArgPheThrGluAsnPheAlaAlaLysAlaGlyValAlaValGlyThrSerSerGlySer 352
Qy 836 GGCATGCTTATGGTCTCTCTGTGGTTATGAATTC 871
Db 353 SerAlaAlaTyrHisValGlyValAsnTyrGluTrp 364

RESULT 2
S04912
yopA protein - Yersinia enterocolitica plasmid pYV6471/76
C:Species: Yersinia enterocolitica
C>Date: 19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change 08-Oct-1999
C:Accession: S04912
R:Skurnik, M.; Wolf-Watz, H.
Mol. Microbiol. 3, 517-529, 1989
A:Title: Analysis of the yopA gene encoding the YopI virulence determinants of Yersinia
A:Reference number: S04910; MUID:89343638; PMID:2761389
A:Accession: S04912
A:Molecule type: DNA
A:Residues: 1-455 <SKU>
A:Cross-references: EMBL:X13882; NID:g48606; PIDN:CAA32086.1; PID:g48607
C:Genetics:
A:Genome: plasmid pYV6471/76

Alignment Scores:
Pred. No.: 5.86e-08 Length: 455
Score: 186.50 Matches: 47
Percent Similarity: 52.17% Conservative: 25
Best Local Similarity: 34.06% Mismatches: 57
Query Match: 9.11% Indels: 9
DB: 2 Gaps: 4

US-10-030-529A-1 (1-1168) x S04912 (1-455)
Qy 458 AATTGTATAGTTATTTTAAACGATTTAAAGACAGGATTTTAAATTAAGATTCTTGATGCA 517
Db 327 AsnValTyrAla-----AspSerLysSerHisThrLeuLysThrAlaAsnSer 343
Qy 518 CGTATTTCCAAAATAAACAATAATGATATCAATAGTAAATATTTACTAGAACGGGT 577
Db 344 TyrThrAspValThrValSerAsn-----SerThrLysLysAlaIleArgGluSerAsn 361
Qy 578 ACTTATTATGATGATCTTATCGTATGATGGAACAAATACACATAATATCAATAGTTG 637
Db 362 GlnTyrThrAspHisLysPheArgGlnLeuAsp-----AsnArgLeuAspLysLeu 378
Qy 638 TCTAAGAATGTCAACTGGTTTACCGCAACCAATCAGCATGCTCTATGTTAGTCACCA 697
Db 379 AspThrArgValAspLysPheGlyLeuAlaSerSerAlaAlaLeuAsnSerLeuPheGlnPro 398
Qy 698 AATGTTGTAGGCAAAACGAGCGCTTCTGCTGGGTAGGAGGTATAGAGATAAAACTGCA 757
Db 399 TyrGlyValGlyLysValAsnPheThrAlaGlyValGlyGlyTyrArgSerSerGlnAla 418
Qy 758 TTAGCCATTGGTGGCTGCATCGCATTTACTGATTCGCTTTACCGCTAAAGCGGGTGTAGCG 817
Db 419 LeuAlaIleGlySerGlyTyrArgValAsnGluAsnValAlaLeuLysAlaGlyValAla 438
```

```
Qy 818 TTCAATACCTACATGCGGCATGCTCTATGTTGGTCTTCTGTTGTTATGAATTC 871
Db 439 Tyr---AlaGlySerSerAspValMetTyrAsnAlaSerPheAsnIleGluTrp 455

RESULT 3
S04911
yopA protein - Yersinia enterocolitica virulence plasmid pYV8081
C:Species: Yersinia enterocolitica
C>Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 08-Oct-1999
C:Accession: S04911
R:Skurnik, M.; Wolf-Watz, H.
Mol. Microbiol. 3, 517-529, 1989
A:Title: Analysis of the yopA gene encoding the YopI virulence determinants of Yersinia
A:Reference number: S04910; MUID:89343638; PMID:2761389
A:Accession: S04911
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-422 <SKU>
A:Cross-references: EMBL:X13881; NID:g48609; PIDN:CAA32085.1; PID:g48610
A:Note: the authors translated the codon CCA for residue 282 as Thr

Alignment Scores:
Pred. No.: 6.45e-08 Length: 422
Score: 186.00 Matches: 44
Percent Similarity: 51.56% Conservative: 22
Best Local Similarity: 34.38% Mismatches: 46
Query Match: 9.08% Indels: 16
DB: 2 Gaps: 3

US-10-030-529A-1 (1-1168) x S04911 (1-422)
Qy 524 TCCAAAATAAACAATAATATTGATCTATTAAGTAATAT----- 562
Db 299 SerLysSerSerHisThrLeuGlnThrAlaAsnSerTyrThrAspValThrValSerAsn 318
Qy 563 -----TTACTAGAACCTGGTACTTATTAGATGATCTTATTCGTATGATG 607
Db 319 SerThrLysLysAlaIleArgGluSerAsnGlnTyrThrAspHisLysPheArgGlnLeu 338
Qy 608 GAACAAAATACACATAATATCAATAGTTGCTTAAAGAAATTCGAACTGGTTAGCCAAC 667
Db 339 Asp-----AsnArgLeuAspLysLeuAspThrArgValAspLysGlyLeuAlaSer 355
Qy 668 CAATCAGCATTTGCTATGTTAGTGCAACCAATCGTGTAGGCAAAACGAGCGTTTCTGCT 727
Db 356 SerAlaAlaLeuAsnSerLeuPheGlnProTyrGlyValGlyLysValAsnPheThrAla 375
Qy 728 GCGTAGGAGGTTATAGAGATAAACTGCATTAGCCATTGGTGTGCGCTCACGCAATCTACT 787
Db 376 GlyValGlyGlyTyrArgSerSerGlnAlaLeuAlaIleGlySerGlyTyrArgValAsn 395
Qy 788 GATCGCTTTACCGTTAAAGCGGTGTAGCGTTCAATACCTACATGCGGCATCTCTAT 847
Db 396 GluSerValAlaLeuLysAlaGlyValAlaTyr---AlaGlySerSerAspValMetTyr 414
Qy 848 GGTGCTTCTGTTGTTATGAATTC 871
Db 415 AsnAlaSerPheAsnIleGluTrp 422

RESULT 4
S04534
invasin precursor - Yersinia pseudotuberculosis plasmid pIB1
C:Species: Yersinia pseudotuberculosis
C>Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 08-Oct-1999
C:Accession: S04534; S04910
R:Rosqvist, R.; Skurnik, M.; Wolf-Watz, H.
Nature 334, 522-525, 1988
A:Title: Increased virulence of Yersinia pseudotuberculosis by two independent mutations
A:Reference number: S04534; MUID:88302441; PMID:3043229
A:Accession: S04534
A>Status: preliminary
A:Molecule type: DNA
```

Pred. No.:	0.000242	Length:	403
Score:	144.00	Matches:	29
Percent Similarity:	81.40%	Conservative:	6
Best Local Similarity:	67.44%	Mismatches:	8
Query Match:	7.22%	Indels:	0
DB:	1	Gaps:	0
US-10-030-529A-1 (1-1168) x H64142 (1-403)			
Qy	1165	CGCACCTTTAACGGCTTAATTTTAGAACATTTTAGAAAAATTCGATAGAGTGATCCCAA	1106
Db	358	ArgThrPheAsnGlyLeuIleLeuGluHisLeuGluGluIleProAspGluGlyThrIle	377
Qy	1105	TTTAGGCTAAACCAACTAATATATACCGTTTTHAGAGTCGCGGATAATATGTTGAAAAAA	1046
Db	378	CysGluIleAspGlyLeuLeuIleThrIleLeuGluValGlyAspAsnMetIleLysGln	397
Qy	1045	GTGNAAGCTA	1037
Db	398	AlaLysVal	400
RESULT 6			
AH3277			
immunoglobulin-binding protein eibE [imported] - Brucella melitensis (strain 161			
C;Species: Brucella melitensis			
C;Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 01-Feb-2002			
C;Accession: AH3277			
R;DeIVecchio, V.G.; Kapatal, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.;			
; Mazur, M.; Goltzman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.			
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002			
A;Title: The genome sequence of the facultative intracellular pathogen Brucella			
A;Reference number: AD3252; PMID:11756688			
A;Accession: AH3277			
A;Status: preliminary			
A;Molecule type: DNA			
A;Residues: 1-155 <KUR>			
A;Cross-references: GB:AE008917; PIDN:AAL51387.1; PID:g17982090; GSPDB:GN00190			
A;Experimental source: strain 16M			
C;Genetics:			
A;Gene: BMIEI0205			
A;Map position: I			
Alignment Scores:			
Pred. No.:	0.00245	Length:	155
Score:	132.00	Matches:	40
Percent Similarity:	45.52%	Conservative:	21
Best Local Similarity:	29.85%	Mismatches:	35
Query Match:	6.45%	Indels:	38
DB:	2	Gaps:	5
US-10-030-529A-1 (1-1168) x AH3277 (1-155)			
Qy	521	ATTTCACAAATTAACAAATATTGTACTATTAAGTAATATTTACTAGACTGGTACT	580
Db	32	IleGlyClnAsnArgThrSerIleSerThrAsnAlaLysGlyValala	47
Qy	581	TATTAGATGATCTTATCGTATGATGAAGCAAAATACA	619
Db	48	-----AspAsnArgAlaAlaIleArgGlnAsnSerAlaAlaIleSerAlaLeuGly	64
Qy	620	-----CATATATCAATAAGTTCTCTAAAGAAATTGCAAACTGGT	658
Db	65	GlnArgValAspGlyLeuGlnGlyGlnIleAsnSerAlaArgLysGluAlaArgAlaGly	84
Qy	659	TTTAGCAACCAATCAGCATTGCTATGTAGTGCACACCAAAATGGTGTAGGCAAAACGAC	718
Db	85	AlaAlaAsnAlaAlaAlaLeuSerGlyLeuArgTyrAspAsnArgProGlyLysValSer	104
Qy	719	GTUUTCTGCTCCGTAGAGGTTATAGAGATAAACTGCATTAGCCATTGCTGTCGCTCA	778
Db	105	IleAlaThrGlyValGlyGlyPheLysGlySerThrAlaLeuAlaGlyIleGly	123
Qy	779	CGCATTACTGATCGCTTTTACCGCTAAA-----CGGGGTGTAGCG	817

Nature 413, 848-852, 2001
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
A:Title: Complete genome sequence of a multiple drug resistant *Salmonella enterica* serov
A:Reference number: AB0502; MUID:21534947; PMID:11677608

A:Accession: AC0834
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-413 <PAR>
A:Cross-references: GB:AL513382; PIDN:CAD05858.1; PID:gl6503833; GSPDB:GN00176
C:Genetics:
A:Gene: corB
C:Superfamily: hypothetical protein HI0107

Alignment Scores:
Pred. No.: 0.00459 Length: 413
Score: 129.00 Matches: 24
Percent Similarity: 81.40% Conservative: 11
Best Local Similarity: 55.81% Mismatches: 8
Query Match: 6.47% Indels: 0
DB: 2 Gaps: 0

US-10-030-529A-1 (1-1168) x AC0834 (1-413)

Qy 1165 CGCACCTTAACGGCTTAATTTAGAACATTTAGAAAAATTCGATAGAAAGTACCCAA 1106
Db 359 ArgThrValAsnGlyValIleLeuGluAlaLeuGluGluIleProValAlaGlyThrArg 378
Qy 1105 TTTAGGCTTAACCACTTAATTAATACCGTTTGTAGAGTGGCGGATATATGTTGAAAAA 1046
Db 379 ValArgIleGluGlnTyrAspIleAspIleValGlnGluAsnMetIleLeuGln 398
Qy 1045 GTGAAGTA 1037
Db 399 ValIysVal 401

RESULT 11

AH0110
probable surface protein (partial) YPO0902 [imported] - *Yersinia pestis* (strain CO92)
C:Species: *Yersinia pestis*
C>Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 02-Nov-2001
C:Accession: AH0110
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.;
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell,
Nature 413, 523-527, 2001
A:Title: Genome sequence of *Yersinia pestis*, the causative agent of plague.
A:Reference number: AB0001; MUID:21470413; PMID:11586360
A:Accession: AH0110
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-658 <KUR>
A:Cross-references: GB:AL590842; PIDN:CAC89747.1; PID:gl5978974; GSPDB:GN00175
C:Genetics:
A:Gene: YPO0902

Alignment Scores:
Pred. No.: 0.0301 Length: 658
Score: 119.50 Matches: 31
Percent Similarity: 49.56% Conservative: 25
Best Local Similarity: 27.43% Mismatches: 50
Query Match: 5.83% Indels: 7
DB: 2 Gaps: 4

US-10-030-529A-1 (1-1168) x AH0110 (1-658)

Qy 539 AATATTGATCTAAGTAAGTAATATTACTAGAACTGGGTACTTATTAGATGATCTTAT 598
Db 551 AsnPheAspGlnLeuLysSerIleSerAsnGlnThrAsnAlaTyrThrAsnGlnArgTyr 570
Qy 599 CGTATGTGGAACAAATACACATAATCAATCAATGTTGCTAAAGAAATTCGAACTGGT 658
Db 571 SerGluLeuLysGln-----AspLeuArgLysGlnAsnSerValLeuSerAlaGly 587

Qy 659 TTAGCAACCAATCAGCATTCGTCTATGTTAGTCAACCAATGGTGTAGCAAAACGAGC 718
Db 588 IleAlaSerAlaMetSerMetAlaSerLeuThrGlnProTyrThrSerGlySerSerMet 607
Qy 719 GTTTCCTGCTGCGTAGGAGGTTATAGAGATAAAACTGCATTAGCCATTGGTGTGCGCTCA 778
Db 608 ThrThrIleGlyAlaAlaSerTyrArgGlyGlnSerAlaLeuSerLeuGlyValSerSer 627
Qy 779 CGCATTAATGAT-----CGCTTTACCGCTAAAGCGGTGTAGCGTTCAATACCTACAAT 832
Db 628 ---IleSerAspSerGlyArgTyrValSerLysLeuGlnAlaSerSerAsnThr---Gln 645
Qy 833 GCGCGCATGCTTATGCTGCTCTGTTGGTTATGAATTC 871
Db 646 GlyAspPheGlyIleGlyValGlyValGlyTyrGlnTrp 658

RESULT 12

AE0169
probable exported protein YPO1387 [imported] - *Yersinia pestis* (strain CO92)
C:Species: *Yersinia pestis*
C>Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 02-Nov-2001
C:Accession: AE0169
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.;
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell,
Nature 413, 523-527, 2001
A:Title: Genome sequence of *Yersinia pestis*, the causative agent of plague.
A:Reference number: AB0001; MUID:21470413; PMID:11586360
A:Accession: AE0169
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-364 <KUR>
A:Cross-references: GB:AL590842; PIDN:CAC90216.1; PID:gl5979436; GSPDB:GN00175
C:Genetics:
A:Gene: YPO1387

Alignment Scores:

Pred. No.: 0.116 Length: 364
Score: 112.50 Matches: 40
Percent Similarity: 43.62% Conservative: 25
Best Local Similarity: 26.85% Mismatches: 69
Query Match: 5.45% Indels: 15
DB: 2 Gaps: 6

US-10-030-529A-1 (1-1168) x AE0169 (1-364)

Qy 431 GGAATAATCGCAGCAGCTGCAAAATTCGAATTTGATAGTATTATTAAACGATTAAAGACAC 490
Db 226 GlyLysSerLysGluThrLeuAsn---AsnThrTyrAspTyrVal----- 239
Qy 491 GATTTTAAATTAATA-----GTTCTTGATGCACGCTATTTCACAAAATAAAACAAATATT 544
Db 240 AspSerLysValSerSerIleValTyrAspValAsnSerTyrThrAspLysThrValAsn 259
Qy 545 GATACTATAAGTAATATTTACTAGAACTGGGTACTTATTAGATGATCTTATCGTATG 604
Db 260 ThrAlaPheGluThrSerLeuSerAspAlaLysSerTyrValAspAspLysTyr----- 277
Qy 605 ATGGAACAAATACACATAATATCAATAAGTTGCTAAAGAAATTCGAACTGGTTAGCC 664
Db 278 ---AsnGlnLeuSerAspLysValAsnLysAsnPheAsnLysThrAsnAlaGlyIleSer 296
Qy 665 AACCAATCAGCATTCGTCTATGTTAGTCAACCAATGGTGTAGCAAAACGAGCGTTTCT 724
Db 297 GlyAlaMetAlaMetSerGlyIleProGlnLysPheGlyTyrGluLys---SerPheGly 315
Qy 725 GCTGCGGTAGAGGTTATAGAGATAAACTGCATTAGCCATTGGTGTGCGCTCAGGCATT 784
Db 316 MetAlaIleGlyAlaTyrArgGlyGlnSerAlaLeuAlaValGlyGlyAspTyrAsnIle 335
Qy 785 ACTGATCGCTTACCGCTAAAGCGGTGTAGCGTTCAATACCTACAATGGCGGCTGTCT 844
Db 336 AsnHisLysThrIleThrArgValAsnValSerAlaAspThrGluGlyGly----- 352

QY 845 TATGCTGCTCTCTGTTGGTTATGAATTC 871
 |||||
 ::
 Db 353 ValGlyValAlaAlaGlyPheAlaPhe 361
 |||||
 ::
 RESULT 13
 D90697
 adhesin/invasin-like protein [imported] - Escherichia coli (strain O157:H7, substrain R1
 C;Species: Escherichia coli
 C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
 C;Accession: D90697
 R;Hayashi, T.; Makino, K.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
 gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
 DNA Res. 8, 11-22, 2001
 A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and geno
 A;Reference number: A99629; MUID:21156231; PMID:11258796
 A;Accession: D90697
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-338 <DNA>
 A;Cross-references: GB:BA000007; PIDN:BA033971.1; PID:gl13360006; GSPDB:GN00154
 A;Experimental source: strain O157:H7, substrain R1MD 0509952
 C;Genetics:
 A;Gene: EC0548

Alignment Scores:
 Pred. No.: 0.208 Length: 338
 Score: 109.50 Matches: 56
 Percent Similarity: 31.54% Conservatives: 32
 Best Local Similarity: 20.07% Mismatches: 92
 Query Match: 5.35% Indels: 99
 DB: Gaps: 2

US-10-030-529A-1 (1-1168) x D90697 (1-338)

QY 317 ACTTATCTTGAATTACAGCATTATATGCTTATATCTCTGTTCTCGTCACATATGCTCT 376
 |||||
 |||||
 Db 65 ThrTyrLeuThrGluHisHisTyrIleProSerGluThrProAspThrThrGlnThrPro 84
 |||||
 |||||
 QY 377 GCGGTTCTCTAGCCCT----- 394
 |||||
 |||||
 Db 85 ThrValGlnThrAspProAspAlaGlyGlnLysThrValAlaAlaThrGlyAspValGln 104
 |||||
 |||||
 QY 395 -----ATACGTTTATATCCGATGCTGATCCTGAT 424
 |||||
 |||||
 Db 105 ThrThrAlaArgTyrGlnSerMetIleAsnAlaArgGlnSerAlaValThrAspValGln 124
 |||||
 |||||
 QY 425 CAACCTGGAATAAATCGGCAGCAGCTGAAATTG----- 457
 |||||
 |||||
 Db 125 GlnThrGlnIleThrGluGlnGlnAlaGlnIleValAlaThrGlnLysThrLeuAlaAla 144
 |||||
 |||||
 QY 458 -----AATTGTATAGTTAT----- 472
 |||||
 |||||
 Db 145 ThrGlyAspThrGlnAsnThrAlaHisTyrGlnGluMetIleAsnAlaArgLeuAlaAla 164
 |||||
 |||||
 QY 473 -----TTTAAACGATTAAAGA 487
 |||||
 |||||
 Db 165 GlnAsnGluAlaAsnGlnArgThrAlaThrGluGlnGlyGlnLysMetAsnAlaLeuThr 184
 |||||
 |||||
 QY 629 AATAAGTTGTCT-----AAAGAA 646
 |||||
 |||||

Db 245 AsnThrLeuAsnAsnHisPheSerSerLeuLysAsnGluValAlaAspAsnArgLysGlu 264
 QY 647 TTGCAAACTGGTTAGCAACCAATCAGCATTGCTATGTAGTGCACCAACCAATGGTGA 706
 |||||
 |||||
 Db 265 AlaAsnAlaGlyThrAla-----SerAlaIleAlaIleAlaSerGlnProGln----- 280
 |||||
 |||||
 QY 707 GGC AAAACGAGC-----GTTTCTGCTCGGTAGGAGGTATAGAGATAAAATC 754
 |||||
 |||||
 Db 281 VallysThrGlyAspValMetMetValSerAlaGlyAlaGlyThrPheAsnGlyGluSer 300
 |||||
 |||||
 QY 755 GCATTAGCCATTGGTTCGGCTCAGCATTAATCGCTTACCGCTAAAGCGGTGTA 814
 |||||
 |||||
 Db 301 AlaValSerValGlyThrSerPheAsnAlaGlyThrHisThrValLeuLysAlaGlyIle 320
 |||||
 |||||
 QY 815 GCGTTCAATCACTACAATGGCGCATGCTTATGCTGCTCTGTTGGTTATGAATTC 871
 |||||
 |||||
 Db 321 SerAlaAspThr---GlnSerAspPheGlyAlaGlyValGlyValGlyTyrSerPhe 338
 |||||
 |||||
 RESULT 14
 G85547
 adhesin/invasin-like protein [similarity] - Escherichia coli (strain O157:H7, substrain
 C;Species: Escherichia coli
 C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 02-Nov-2001
 C;Accession: G85547
 R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
 iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
 Nature 409, 529-533, 2001
 A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
 A;Reference number: A85480; MUID:21074935; PMID:11206551
 A;Accession: G85547
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-338 <STO>
 A;Cross-references: GB:AE005174; NID:gl2513376; PIDN:AAG54843.1; GSPDB:GN00145; UNGP:206
 A;Experimental source: strain O157:H7, substrain EDL933
 C;Genetics:
 A;Gene: Z0639

Alignment Scores:
 Pred. No.: 0.208 Length: 338
 Score: 109.50 Matches: 56
 Percent Similarity: 31.54% Conservatives: 32
 Best Local Similarity: 20.07% Mismatches: 92
 Query Match: 5.35% Indels: 99
 DB: Gaps: 2

US-10-030-529A-1 (1-1168) x G85547 (1-338)

QY 317 ACTTATCTTGAATTACAGCATTATATGCTTATATCTCTGTTCTCGTCACATATGCTCT 376
 |||||
 |||||
 Db 65 ThrTyrLeuThrGluHisHisTyrIleProSerGluThrProAspThrThrGlnThrPro 84
 |||||
 |||||
 QY 377 GCGGTTCTCTAGCCCT----- 394
 |||||
 |||||
 Db 85 ThrValGlnThrAspProAspAlaGlyGlnLysThrValAlaAlaThrGlyAspValGln 104
 |||||
 |||||
 QY 395 -----ATACGTTTATATCCGATGCTGATCCTGAT 424
 |||||
 |||||
 Db 105 ThrThrAlaArgTyrGlnSerMetIleAsnAlaArgGlnSerAlaValThrAspValGln 124
 |||||
 |||||
 QY 425 CAACCTGGAATAAATCGGCAGCAGCTGAAATTG----- 457
 |||||
 |||||
 Db 125 GlnThrGlnIleThrGluGlnGlnAlaGlnIleValAlaThrGlnLysThrLeuAlaAla 144
 |||||
 |||||
 QY 458 -----AATTGTATAGTTAT----- 472
 |||||
 |||||
 Db 145 ThrGlyAspThrGlnAsnThrAlaHisTyrGlnGluMetIleAsnAlaArgLeuAlaAla 164
 |||||
 |||||
 QY 473 -----TTTAAACGATTAAAGA 487
 |||||
 |||||
 Db 165 GlnAsnGluAlaAsnGlnArgThrAlaThrGluGlnGlyGlnLysMetAsnAlaLeuThr 184
 |||||
 |||||
 QY 488 CACGATTTTAAATTAAGTTCTTTCGATGACGATTTTCCAAAAATAAACAA----- 538
 |||||
 |||||

Search completed: May 13, 2004, 08:29:37
Job time : 32.5 secs

```
Db      185 ThrAspValAlaValGlnGlnGlnAsnGluArgThrGlnTyrAspLysGlnMetGlnSer 204
Qy      539 -----AATATTGATCTATAAGTAATAATTTACTA 568
Db      205 LeuAlaGlnGluSerAlaGlnAlaHisGlnGlnIleAspSerLeuSerGlnAspValThr 224
Qy      569 GAATGGGTACTTATTAGATGATCTTATCGTATGGAACAATAACACATAATATC 628
Db      225 GlnThrHisGlnGlnLeuThrAsnThrGlnLysArgValAlaAspAsnSerGlnGlnIle 244
Qy      629 AATAAGTTGTCT-----AAAGAA 646
Db      245 AsnThrLeuAsnAsnHisPheSerSerLeuLysAsnGluValAlaAspAsnArgLysGlu 264
Qy      647 TTGCAAACTGGTTAGCAACCAATGACATGCTATGTTAGTGCAACCAATGGTGTA 706
Db      265 AlaAsnAlaGlyThrAla-----SerAlaIleAlaIleAlaSerGlnProGln----- 280
Qy      707 GGCRAAACGAGC-----GTTCTGCTCGGTAGGAGGTTATAGAGATAAAACT 754
Db      281 ValLysThrGlyAspValMetMetValSerAlaGlyAlaGlyThrPheAsnGlyGluSer 300
Qy      755 GCATTAGCATGCTGTCGCTCAGCATCTACTCGCTTACCGCTAAAGCGGTGTA 814
Db      301 AlaValSerValGlyThrSerPheAsnAlaGlyThrHisThrValLeuLysAlaGlyIle 320
Qy      815 GCGTTCAATCACTACATGCGGCGATGCTTATGCTGCTTGTGGTTATGAATTC 871
Db      321 SerAlaAspThr---GlnSerAspPheGlyAlaGlyValGlyValGlyTyrSerPhe 338
```

RESULT 15

AF0400
probable membrane protein YPO3298 [imported] - Yersinia pestis (strain CO92)
C;Species: Yersinia pestis
C;Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 14-Dec-2001
C;Accession: AF0400
R;Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.; deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.; il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell, Nature 413, 523-527, 2001
A;Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A;Reference number: AB0001; PMID:21470413, PMID:11586360
A;Accession: AF0400
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-427 <KUR>
A;Cross-references: GB:AL590842; PIDN:CAC92530.1; PID:gl5981228; GSPDB:GN00175
C;Genetics:
A;Gene: YPO3298
C;Superfamily: hypothetical protein HI0107

Alignment Scores:
Pred. No.: 0.343 Length: 427
Score: 107.00 Matches: 21
Percent Similarity: 67.44% Conservative: 8
Best Local Similarity: 48.84% Mismatches: 14
Query Match: 5.36% Indels: 0
DB: 2 Gaps: 0

US-10-030-529A-1 (1-1168) x AF0400 (1-427)

```
Qy      1165 CGCACCTTTTACGGCTTAATTTTAGAATTTAGAAAAAATTCGATAGAGTACCCAA 1106
Db      374 ArgThrIleAsnGlyMetLeuLeuGluGluGluGluGluIleProGlnAlaHisIleHis 393
Qy      1105 TTTAGGCTAAACCAACTTAATATATACCGTTTATAGAGTGGCGGATATATGCGTAAAAAA 1046
Db      394 ValArgLeuGlyAsnTyrGlnIleGluValLeuAspValGlnGlnAsnMetIleLysArg 413
Qy      1045 GTGAAGTA 1037
Db      414 ValLysIle 416
```

This Page Blank (uspto)